2021

Living in an oil-polluted Region in Nigeria as a Risk Factor in Colorectal Cancer Development

Steve Nwachi Onya
Walden University

Walden University

College of Health Professions

This is to certify that the doctoral dissertation by

Steve N. Onya

has been found to be complete and satisfactory in all respects,
and that any and all revisions required by
the review committee have been made.

Review Committee
Dr. German Gonzalez, Committee Chairperson, Public Health Faculty
Dr. Lee Caplan, Committee Member, Public Health Faculty
Dr. Wen-hung Kuo, University Reviewer, Public Health Faculty

Chief Academic Officer and Provost
Sue Subocz, Ph.D.

Walden University
2021

Abstract

Living in an Oil-Polluted Region in Nigeria as a Risk Factor in Colorectal Cancer Development

by

Steve N. Onya

MSc, Roehampton University, 2017

BPharm, University of Nigeria, Nsukka, 1989

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Public Health

Walden University

November 2021

Abstract

Due to differences in exposure to various risk factors, colorectal cancer (CRC) rates differ in regions of Nigeria. Although the people living in the Southsouth, an oil-producing region, are concerned that they are most affected, there is insufficient research to address these concerns. The purpose of this quantitative ecological study was to investigate whether living in the oil-polluted region in Nigeria is associated with an excess CRC incidence rate. The potential spatial autocorrelation of CRC incidence among the contiguous regions was also examined. The study was grounded in the ecosocial theoretical framework. Data were extracted from the Nigeria National System of Cancer Registries for 2009-2016, online real-time IQAIR data on particulate matter (PM 2.5) exposure, and the 2018 Nigeria Demographic and Health Survey. The data set included 526 CRC cases, representing 298 (56.65%) male and 228 (43.35%) female individuals. One-way ANOVA and ANCOVA were used to assess CRC disparity and adjust for the covariate. The *t*-test and multiple linear regression tests investigated gender as a predictor variable in CRC disparity in this population and adjusted for the various socioeconomic covariates. The study's main findings showed no statistically significant association between living in the oil-polluted region and the risk of excess CRC ($p > 0.05$). The study also showed a weak negative spatial autocorrelation that was not statistically significant ($I = -0.365$; $p > 0.05$). This study may contribute to positive social change by providing evidence-based data on the relationship between environmental exposures and the risk of chronic disease like CRC. This knowledge may inform Pareto efficiencies in public health interventions as well as educate members of the public.

Living in an Oil-Polluted Region in Nigeria as a Risk Factor in Colorectal Cancer Development

by

Steve N. Onya

MSc, Roehampton University, 2017

BPharm, University of Nigeria, Nsukka, 1989

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Public Health

Walden University

February 2022

Dedication

I dedicate this dissertation work to God Almighty, from whom all blessings flow.

Acknowledgments

I want to thank Dr. G. Gonzalez, Dr. L. S. Caplan, and Dr. W. Kuo for their robust support throughout the dissertation period. I would also like to take the opportunity to thank all the faculty members and staff of the College of Health Professions for their invaluable support.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction to the Study

The aim of this study was to examine if living in Nigeria's oil-polluted South South region poses a greater risk of excess colorectal cancer (CRC) incidence when compared to the other regions. According to experts, CRC, which is classified generally as a gastrointestinal carcinoma that usually starts with a generalized disorder of cell replication, is caused by modifiable and nonmodifiable putative risk factors (Ray-Offor et al., 2020), such as environmental pollution, including pollution from crude oil exploitation (Johnston et al., 2019; Kachuri et al., 2016; Vargas et al., 2020). Therefore, the distribution of CRC in Nigeria may reflect the exposure of the six regions in Nigeria to different environmental and other putative risk factors.

Nigeria is a federation with six geographically, culturally, and socioeconomically distinct regions or geopolitical zones, as seen in Figure 1, with crude oil export as the main source of foreign earnings. The Niger Delta region, which loosely consists of nine oil-producing states (Abia, Cross River, Bayelsa, Delta, Imo, Ondo, Edo, Rivers, and Akwa-Ibom), accounts for the majority of Nigeria foreign earnings (Bodo et al., 2019). The higher level of foreign earnings is because of the dominance of crude oil in this region, which makes residents vulnerable to environmental degradation with its health consequences, such as CRC. According to Elum et al. (2016), there is a common belief within the Niger Delta region, which includes the South South region or geopolitical zone, that the government of Nigeria and oil multinationals are not positively disposed to improving the socioeconomic and environmental well-being of oil-host communities, especially in events of oil spillage. The research findings by Osuagwu and Olaifa (2018) and Udotong et al.

(2017) revealed the negative effect of the oil spill on the livelihood and well-being of persons living in oil-producing areas. People living in the oil-producing region, especially the South South region of Nigeria, argue that they are more at risk of developing poor health outcomes, such as CRC, because of pollution from oil and gas exploitation (Osuagwu & Olaifa, 2018), as seen in Figure 2, than those living in other regions with other sources of main pollution, as seen in Figure 3. According to Ebegbulem et al. (2013) and Bodo and Gimah (2019), the South South region generally bears most of the burdens of environmental devastations due to oil and gas exploration by multinational oil companies. Inoni et al. (2006) also argued that a 10% increase in oil spills reduces crop yield and farm income by about 1.3% and 5%. Communities in the South South region have resorted to self-determination struggle and other agitation forms to protest their marginalization and exposure to risk factors of poor health outcomes.

However, it is important to note that people living in non-oil-producing or -polluted regions in Nigeria also develop CRC. Although environmental pollution, including pollution from oil (Kachuri et al., 2016; Johnston et al., 2019; Vargas et al., 2020), is associated with CRC, there is evidence linking CRC to both modifiable and nonmodifiable risk factors (Ray-Offor et al., 2020). These risk factors include, but are not limited to, various socioenvironmental factors, such as socioeconomic status (Hurtado et al., 2015), built-environmental factors (Canchola et al., 2017), gene-environmental interaction (Gabriel et al., 2018), factors related to lifestyle (Chao et al., 2000; Kunzmann et al., 2015; Rossi et al., 2018; Slattery et al., 2003), barriers to CRC screening (Azeem et al.,2016), maternal stress risk factors (Bishedsari et al., 2014), and the life-course related events

(Clarke & Joshua, 2017). Therefore, CRC may be disproportionately distributed in Nigeria across the regions that reflect their exposure to the various putative risk factors, including living in an oil-polluted environment. Addressing CRC disparity in these populations may promote positive social change. Specifically, data from this study may compel policy makers to institute the necessary institutional reforms and behavioral changes that motivate social change across the country. According to de la Salonniere (2017), social change encompasses the transformation of behavior, culture, social institutions, and social structure over time.

CRC is generally classified as a gastrointestinal carcinoma. It usually starts with a generalized disorder of cell replication with the appearance of clusters of enlarged crypts (aberrant crypts) showing proliferative, biochemical, and biomolecular abnormalities. According to Gramados-Romeo (2017), CRC usually starts with a polyp in the intestinal mucosa. However, it can also manifest as an initial benign lesion with the ability to transform into a full malignant lesion depending on its histopathological presentations. Badmos et al. (2018), in their research on the histopathology of CRC, argued that most colorectal malignancies develop from adenomatous polyps. These premalignant adenomatous polyps are usually asymptomatic, making their early detection difficult until patients present themselves at about 50 years. In developing countries like Africa, most diagnoses of CRC (57%) occurs at the late Duke's Stage B (Vega et al., 2015). Therefore, CRC is intestinal cancer that starts as an asymptomatic benign polyp, which, if undetected, will progress to colorectal malignancies.

The distribution of CRC across the six regions in Nigeria is influenced by the prevalence of the putative risk factors and the spatial autocorrelation between the contiguous regions. The first law of geography posits that "everything is related to everything else, but near things are more related than distant things'' (Tobler, 1970, as cited in Ahmadi and Al-Zahrani, 2013, p. 7210). Therefore, spatial autocorrelation is when the value of a variable at a specified geographic location depends on its values at adjacent locations (Rosenberg, 2000).

This study is important because the evidence may inform the creation of appropriate interventions by the Nigerian government to address social problems arising from the perceived CRC disparity across the regions. Addressing the spatial autocorrelation of CRC between contiguous regions may enable public health officers to make an informed decision regarding CRC clustering, dispersion, or random occurrence in the regions. This study fills the gap in knowledge and literature and contributes to contemporary chronic disease epidemiology, including the life-course approach to disease prevention. The evidence from this study may support positive health behavior and other institutional reforms that will motivate social change across the country.

Epidemiology of Colorectal Cancer

Globally, CRC is the third most common cancer that accounted for over 700,000 deaths annually on the average, as reported in the extant literature, with an estimated incidence rate of 2.2 million cases in the next 10 years (Bray et al., 2018). In the European continent, most studies have reported that the crude incidence rate of CRC is about 60.30 per 100,000 population (Bray et al. (2018). The male population in Europe accounts for

about 67.60 per 100,000, while the female population incidence is about 53.50 per 100,000 standard world population (Bray et al. (2018). According to Irabor et al. (2014), the crude incidence rate of CRC for developing countries, especially South Saharan Africa, is 4.04 per 100,000 population. In the South Saharan Africa countries, the male population accounts for 4.38 incidents per 100,000 population, while rate in the female population is 3.69 incidents per 100,000. In Africa, the incidence rate increases with age, peaking at 75 years. These statistics show that developed countries have a bigger CRC burden than developing countries.

CRC incidence in developing countries is much lower than the developed countries (Gaylard & Ramos, 2018). Nigeria's incidence of CRC is considered low at 3.4 per 100,000 population, according to Irabor et al. (2014). However, the emerging extant pieces of literature have shown that the incidence of CRC has increased in Africa, especially sub-Saharan African countries such as Nigeria, where the incidence was considered low in previous research reports (Rahman, 2010). Irabor (2018), in a review of articles, concluded that researchers must accept the inevitable that CRC increase has come to stay in the West African region. It is also instructive to note that in developing countries like Nigeria, where oil export is their mainstay, findings from a large epidemiological study show a link between gases associated with oil pollution such as particulate matter (PM 2.5) and NO2 with kidney, bladder, and CRC death (Turner et al., 2018). Moreover, other developing countries outside Africa, like those in Western Asia, which previously and historically reported a low incidence of CRC, are beginning to report an increase in new cases and prevalence of CRC in their populations (Al-Jaberi, 1999). In addition, Basaleem and Al-

Sakkaf (2004) reported the rising incidence of CRC in countries like Yemen, Jordan, and Egypt. The increasing rate of CRC in developing and low-resource countries is a departure from the evidence of previous studies (Colorectal Cancer Collaborators. 2019). In summary, the CRC incidence rate that was relatively low in developing countries like Nigeria has shown an upward trend due to western lifestyle, environmental pollution, socioeconomic status, built environment, and other risk factors associated with oil pollution (Johnston et al., 2019).

In this chapter, I address the background of the study and present the problem statement, research purpose, research questions (RQs) and hypotheses, theoretical framework, nature of the study, and key definitions. I also discuss the assumptions, scope and delimitations, limitations, and significance of the study, including its implications for positive social change. The chapter concludes with a summary and transition to Chapter 2.

Background

Several pieces of extant evidence link oil pollution with poor health outcomes, including different types of cancer. Johnston et al. (2019) provided evidence that suggested a positive relationship between the impacts of oil spill exposure and cancer. Chinedu and Chukwuemeka (2018) argued that heavy metals from crude oil spills are risk factors in poor health outcomes. Long-term adverse health effects have also been linked to the exposures to oil drilling activities (D'Andrea & Reddy, 2018). There is also evidence suggesting that there are other risk factors associated with CRC, such as environmental air pollution and other socioeconomic and life-course related risk factors. For instance, a landmark epidemiological study has shown that environmental pollutants like N02 and PM

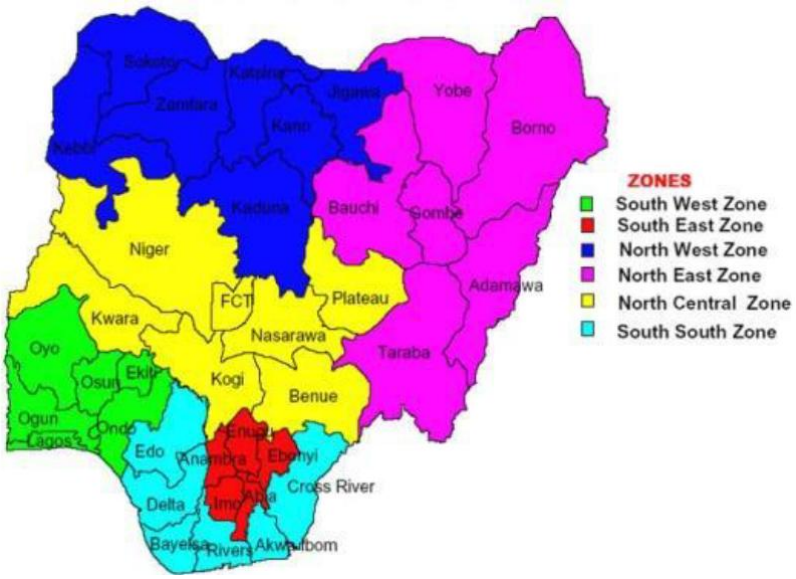
2.5 were positively associated with CRC development (Turner et al., 2018). Therefore, using a life-course approach and theoretical framework may provide better evidence to link CRC with the cumulative effects of several risk factors over the individual life-course. An ecological study is appropriate to study CRC because differences in environmental exposures are usually bigger within groups than at the individual levels (Stevenson & McClure, 2005).

In Nigeria, about 244 000 barrels of oil are spilled every year in the Niger Delta region, which includes the South South region or geopolitical zone (Ordinioha & Brisibe, 2013). The South South region thus bears the burden of oil spillage and its health consequences in Nigeria. This region is responsible for over 85% of Nigeria's total foreign earnings from oil export (Bodo et al., 2019). The negative effect of oil spill on crop production, as shown in Figure 2, has resulted in reduced crop yield, land productivity, and depressed farm income as a 10% increase in oil spill reduced crop yield by 1.3% while farm income plummeted by 5% (Inoni et al., 2006). Communities in this region have resorted to self-determination struggle and other forms of agitation to protest their marginalization and exposure to risk factors of poor health outcomes. However, other regions or zones in Nigeria are also exposed to other forms of pollutants, as shown in Figure 3, that predispose them to various forms of cancer, which may include CRC, and other related poor health outcomes. The foregoing may explain the possibility of some communities outside the oil-producing region having a higher CRC incidence than those living in the oil-producing communities or region.

For instance, Irabor et al. (2014) reported that in Ibadan, a city in Southwest Nigeria, which is a non-oil-producing region, the total CRC prevalence as of 2017 was 120. The incidence rate was about 3.5 per 100,000 population with a mean age of 51 years. Those below 20 years accounted for about 3.3%, with a male: female ratio of 3:1. Those aged 20 to 30 (13.3%) had a sex ratio of 1: 1.6. Those aged between 30 and 40 (31%) had a sex ratio of 1: 1.5. Those aged above 40 (52.4%) had a sex ratio of 1:1. However, in some states of the South South (oil-producing region) and the Southeast contiguous or neighboring region, the CRC incidence rates were lower than those of the Southwest regions within the same age cohort and gender (Ray-Ofor et al., 2020). Also, a 5-year retrospective study in Nigeria of different cancer epidemiology found no difference in cancer incidence between regions or ethnic delineations (Obiorah & Nwafor, 2019).

Figure 1

Map of Nigeria Showing the Six Regions or Zones



Note. From the Federal Ministry of Health Nigeria. National Malaria Elimination Program (https://health.gov.ng/doc/NMEP-Strategic-Plan.pdf). In the public domain.

Figure 2

Oil Pollution from oil spill in the South South Zone

Note. From ‘’Nigeria lost N898.93bn to pipeline vandalization, theft in nine months’’. Online Punch Newspaper. (https://punchng.com/nigeria-lost-n898-93bn-to-pipeline-vandalism-theft-in-nine-months/). In the public domain.

Figure 3

Air Pollution in both the oil and non-oil producing regions.

Note. From ‘’Air Pollution in the Niger Delta Area: Scope, Challenges and Remedies’’. IntechOpen, by Godson Rowland Ana, 2011, IntechOpen (http://www.intechopen.com/books/the-impact-of-airpollution-on-health-economy-environment-and-agricultural-sources/air-pollution-in-the-niger-delta-area-scopechallenges-and-remedies). Creative Commons Attribution-Non CommercialShareAlike-3.0 License.

I selected CRC in this study because, among all the types of cancer associated with oil and other environmental pollution, it is more likely to result from consequences of environmental pollution and degradation such as loss of space for farming, lack of opportunity for physical activities, negative behavioral lifestyle, low intake of fruits and vegetables, and Type 2 diabetes (Park et al., 2007; Slattery et al., 2004). In addition, gases from oil pollution, such as NO2 and PM 2.5, have been found to be positively associated more with CRC than other cancer types (Turner et al., 2017). Other risk factors, such as alcohol consumption and smoking (Marley & Nan, 2016), are all linked to CRC because they are behaviors that are targeted at reducing the stress associated with inequality, poverty-induced social disparities, and socioenvironmental factors in low-resource populations.

To my knowledge, no study had focused on the relationship between these putative risk factors and the development of CRC in the population studied, including those using the ecological study approach. To address this gap in the literature, I used a quantitative, correlational ecological study design to show CRC outcomes as they relate to peculiar environmental hazards in various regions or zones. I used the ecological study design

because differences in exposure between groups are usually bigger than measurement at the individual level and because ecological studies are better at informing policy changes at the population level (Stevenson & McClure, 2005). Ecological studies, especially those involving population-based registries, are useful in controlling for the subjects' temporal and spatial mobility and the spatial scale of exposure assessment that could lead to misclassification in the cumulative exposure estimation (Wahida et al., 2016). In answering the RQs, I considered CRC outcomes a life-long cumulative exposure (Cavelin 2019), including all the exposomes at the various regions that may have been responsible for the current disparity (if any) in CRC incidence. In conclusion, this study's supporting literature suggests a disparity in the burden of CRC across the six geopolitical and ethnically delineated zones in Nigeria. Therefore, the quantitative correlational ecological study design was appropriate to inform differential regional exposure and outcomes, including the opportunity for the appropriate public health interventions to motivate social change in these populations.

Problem Statement

In Nigeria, no study to my knowledge had provided evidence of the relationship between living in an oil-polluted environment and the risk of developing excess CRC in the population, which is important in addressing the concerns of those living in the oil-producing regions. I sought to address this gap in the literature because this knowledge is vital in implementing the appropriate interventions to ensure health equity and positive social change in the population under study. This study is also significant because addressing perceived disparities in health outcomes, such as CRC due to differential

exposure to environmental pollution, may minimize social problems that may affect a population's health (Moses & Olaniyi, 2017).

Several pieces of extant literature provide evidence that link oil pollution with poor health outcomes, such as CRC. Johnston et al. (2019) suggested a positive relationship between the effect of oil spill exposure and various cancer sites. Chinedu and Chukwuemeka (2018) argued that heavy metals from crude oil spills are known risk factors for poor health outcomes, while in a landmark epidemiological study, environmental pollutants like N02 and PM 2.5 were shown to be positively associated with CRC development (Turner et al., 2017). Furthermore, long-term adverse health effects, such as CRC, have been linked to exposure to oil drilling activities (D'Andrea & Reddy, 2018). The evidence from these studies is reinforced by the findings of Vargas (2020) that confirmed the positive association between public health issues, such as CRC, and crude oil production in the Ecuadorian oil-producing territory. However, despite the evidence provided by several studies on the association between oil pollution and CRC, including ecological studies of other populations (Lopez-Abente et al., 2012), no study has so far provided evidence on the association between developing CRC and living in an oil-producing region in Nigeria, including those using quantitative correlational ecological study designs and spatial autocorrelation analysis.

Most researchers studying CRC in other populations have used mainly quantitative, prospective, and retrospective study designs (e.g., Lopez-Abente et al., 2012; Turner et al.,2017). The use of individual data in most of the reviewed literature did not present the opportunity to address exposures or correlates that cannot be easily measured at the

individual level biomarkers. The absence of the use of cumulative exposures as the independent variable in the reviewed literature, including studies conducted in Nigeria, is of methodological and other significance (Miles, 2017). I addressed this gap in the literature by conducting a quantitative correlational ecological study. Therefore, the purpose of this study was to investigate whether exposure to oil pollution from living in oil-producing regions is responsible for any excess CRC outcome compared to living in non-oil-producing regions in Nigeria. I also investigated the confounding impact of spatial autocorrelation of CRC between contiguous regions in Nigeria and the possible gender CRC disparity in the population under study.

Purpose of the Study

The purpose of this quantitative, correlational ecological study was to investigate whether exposure to oil pollution or living near oil-producing areas is associated with any excess CRC outcome across the regions in Nigeria. Specifically, I sought to answer the question of whether living in an oil-polluted region posed a greater risk to CRC incidence than living in other regions in Nigeria with exposures to other forms of pollution than oil. I also investigated the confounding impact of spatial autocorrelation of CRC between contiguous regions in Nigeria. The possible relationship between gender and CRC disparity as a potential measure of occupational hazards and risks within the population under study (Scarselli et al., 2018) was also examined. I based the study purpose on the problem statement in the previous section. Creswell and Creswell (2018) argued that a problem statement helps narrow down the purpose statement in predicting what should be learned in each study. Data from this study are expected to provide evidence for government policy

in the ongoing scholarly inquiry for the differential public health interventions targeted at communities based on their level of exposure to environmental pollution. Data from the study may support public health interventions that reflect differential exposure of the various regions in Nigeria to different CRC putative risk factors. These interventions may minimize any disparity in CRC outcome in the population under study, which may eventually promote health equity and positive social change.

In this study, the dependent variable was CRC incidence across the six independent regions, while the independent variables included regions and gender. I also assessed the impact of covariates like income inequality' (Gini Index), education, wealth quantiles, and access to health insurance. These covariates were important to quantitatively account for their confounding effect in the relationship between the dependent and independent variables in answering the RQs.

Research Questions and Hypotheses

RQ1: What is the relationship between living in the oil-polluted environment or region and CRC incidence in Nigeria?

$H_0$1: There is no statistically significant relationship between living in oil-polluted environment or region and CRC incidence in Nigeria

H_A1*:* There is a statistically significant relationship between living in oil-polluted environment or region and CRC incidence in Nigeria.

The focus of RQ1 was on determining whether there was a statistically significant relationship between living in an oil-polluted region and CRC incidence in Nigeria. The independent variable, geography, was represented by the six independent regions, while

the dependent variable was age-standardized CRC rates. The dependent variable was a continuous variable with a scale level of measurement, while the independent variable was a categorical variable with a nominal level of measurement.

RQ2: What is the relationship between the contiguity of the six regions in Nigeria and the spatial positive autocorrelation of CRC incidence?

$H_0$2: There is no statistically significant relationship between contiguity of the six regions in Nigeria and the spatial positive autocorrelation of CRC incidence,

H_A2: There is a statistically significant relationship between contiguity of the six regions in Nigeria and the spatial positive autocorrelation of CRC incidence.

The focus of RQ2 was on determining whether there was a statistically significant positive spatial autocorrelation of CRC among the six regions in Nigeria. The independent variable was geography, as represented by the contiguous regions, while the dependent variable was CRC incidence rates. The independent variable was a categorical variable with six independent levels and a nominal level of measurement, while the dependent variable was a continuous variable with a scale level of measurement.

RQ3: What is the relationship between gender and CRC incidence within and between Nigeria's various regions after adjusting for covariates, such as education, wealth, Gini Index, and access to health insurance?

$H_0$3: There is no statistically significant relationship between gender and CRC incidence within and between Nigeria's various regions after adjusting for covariates, such as education, wealth, Gini Index, and access to health insurance.

H_A3: There is a statistically significant relationship between gender and CRC incidence within and between Nigeria's various regions after adjusting for covariates, such as education, wealth, Gini Index, and access to health insurance.

The focus of RQ3 was on determining whether there is a statistically significant relationship between gender and CRC incidence in Nigeria. The independent variable was gender while the dependent variable was CRC incidence rates. The independent variable was a dichotomous categorical variable with a nominal level of measurement while the dependent variable was a continuous variable with a scale level of measurement.

Theoretical Framework

I grounded this study in theory based on the ecosocial epidemiological theoretical framework. An ecosocial theory, like other epidemiological theories for disease distribution, is proposed by Nancy Krieger. It is an emerging multilevel theory of disease distribution that integrates social and biological reasonings. It integrates the dynamic, historical, socio-environmental, and ecological perspectives of disease origin to inform disease distribution within a population, including the social inequalities in health. It tends to answer how socioenvironmental and biological factors interact to generate health and disease disparities within and between populations. According to Krieger (2011), ecosocial theory enabled the Faber and Harvard Cancer center working group to identify gaps in knowledge about cancer inequality by addressing the social inequality at multiple levels. I also leveraged other studies, such as those implicitly grounded in ecosocial, psychosocial, and socio-political theories, to inform the theoretical foundation for this study. Some of the reviewed pieces of literature anchored their theoretical argument on ecosocial and

psychosocial theories, which imply that chronic disease risks are because of physical and social exposures across the life course (Ben-Shiomog & Kuh, 2002). Ben-Shiomog and Kuh (2002) argument resonates well with the ecosocial epidemiological theoretical framework to propose an assumption that there is an integration of biological and psychosocial pathways that informs poor health outcomes over time in the population under study. Canchola et al. (2017), Kichuchi et al. (2017), and Zhang et al. (2018), all implicitly used the ecosocial theoretical framework to explain the relationship between social-environmental factors, such as living in poor built neighborhoods, stress, low socioeconomic status, and the risk of developing CRC in some populations.

The socio epidemiologic alternative theories of disease distribution, such as the ecosocial theory, including the sociopolitical and psychosocial theories, recognize that illness or disease is socially produced (Conrad & Parker, 2010). The three theories are premised on the fact that health and disease distribution in the human population occurs in a social context and that this distribution of disease is influenced by social, environmental, economic, political, cultural, and technological features (Krieger, 2011). Specifically, the ecosocial epidemiological theory holds that population distribution of disease, disability, death, and health, and their determinants and deterrents across time and space are influenced by socio-environmental factors (Krieger, 2001). The ecosocial constructs allowed me to deal with pathways and cumulative exposure (Krieger et al., 2018). I leveraged the ecosocial theory to support my independent variable as cumulative exposures beyond the individual environmental pollutants, while the dependent variable is the colorectal cancer incidence cases. The preceding is because the ‘’ill concept,’’ defined by

social epidemiologic theories, is related to the social conditions in which people are born, live, work, and die (Marmot, 2014). The ecosocial theory of disease distribution in this study allowed me to explain the reasons for a possible spatial autocorrelation of CRC between contiguous regions in Nigeria. In the main, the ecosocial epidemiological theoretical foundation allowed this research to treat cumulative exposures in each of the six regions in Nigeria as the independent or exposure variable, including spatial autocorrelation analysis to inform CRC distribution in the populations studied. The preceding is consistent with the research question, research problem, and the proposed ecological study and spatial terms analysis.

Nature of the Study

The nature of this study was a quantitative, correlational ecological method. The quantitative ecological study design assumed that differences in exposure between groups are usually more significant than measurement at the individual level, and ecological studies are better to inform policy changes at the population level (Stevenson & McClure, 2015). The quantitative approach enabled the study to use deductive reasoning to test the hypothesis of the relationship between living in the oil-polluted area and the risk of colorectal cancer development in the population of interest. The quantitative study method also allowed for the positivist epistemological stance because this study used an objective rather than a subjective inquiry approach (Easterby-Smith et al., 2012). It is a quantitative approach because conclusions were designed to be drawn from the research question, hypothesis, and the larger meaning of the result (Creswell & Creswell, 2018). The quantitative method involved the selection of a data set derived from a population-based

cancer registry. In this study, the dependent variable in all three research questions is the CRC incidence, a quantitative continuous variable with a scale level of measurement. In research question two, the independent variable is ‘’Geography’’ or ‘’Region,’’ a categorical variable with six independent groups, measured at the nominal level. In research question three, the independent variable is ‘’Gender,’’ a categorical dichotomous variable with a nominal level of measurement. The covariates included in the statistical model in this study were education, Gini income index, wealth, and access to health insurance. All the covariates were available as group data, measured at the scale level. The covariates adjust for any potential confounding at the group level.

I used a secondary data set to answer the main RQ. This data set is from a population-based cancer registry domiciled with the Federal Ministry of Health. It is a document from the Nigerian National System of Cancer Registries (NSCR) established to provide cancer data incidence from the different regions of the country in addition to ensuring the database integrity and validity based on regular data audits at the various regional cancer registries (Jedy-Agba et al., 2016). The NSCR and the Nigerian Demographic and Health Survey (NDHS) data answered the research questions. While the NSCR data set was a source for the CRC incidence across regions, the 2018 NDHS secondary data set was helpful for data on the covariates of interest. These secondary data were analyzed using the SPSS version 27 after the appropriate data set cleaning to ensure no violation of the assumptions of any of the statistical tests. Thus, the theoretical framework, ecological study design, secondary data sets, and the ‘’Analysis of Variance (ANOVA), t-test, Multiple Linear Regression, and Global Moran spatial autocorrelation

statistical tests were aligned to answer the research questions that lent itself to quantitative research design.

The use of group data in this quantitative ecological study is because environmental exposures often do not differ significantly among individuals within one geographical area (Hatch & Thomas,1993). Besides, some of the exposomes associated with colorectal cancer may not have ‘’correlates,’’ especially the complex mixtures and exposures like noise, hence cannot be measured at the individual level biomarkers (Peters et al., 2012). Also, chronic diseases like colorectal cancer caused by environmental pollution is a life course event, as it may occur because of the accumulation of stress due to inequalities, gene-environmental interaction, and other socio-environmental factors related to the fetal or developmental origins of adult disease, a hypothesis proposed by Godfrey and Barker (Calkins & Devaskar, 2011).

Definitions

Colorectal cancer: A cancer that is generally classified as a gastrointestinal carcinoma that usually starts with a generalized disorder of cell replication and with the appearance of clusters of enlarged crypts (aberrant crypts) showing proliferative, biochemical, and biomolecular abnormalities (Badmos et al., 2018).

Ecological study: A study design in which the researcher uses the mean aggregate data of a group instead of the individual data as a measure of the outcome variable (Wakefield, 2007). The exposure variable can be regional, geographical, or different country-level exposure data.

Exposome: The lifelong integration of all exposures that may influence human health**.** In this study, exposome refers to the sum of all socioenvironmental factors that inform CRC incidence in the target population (see Stingone et al., 2017).

Ecosocial theory: An epidemiological theory that integrates the dynamic, historical, socioenvironmental, and ecological perspective of disease origin to inform understanding of disease distribution within a population, including social inequalities in health (Krieger, 2011).

Gini Index: A measure that is used to measure the levels of income inequality within a given population or across regions in a given country (World Bank, 2013).

Niger Delta region: A region in Nigeria comprising nine oil-producing states that make up mainly the South South geopolitical zones. The region is responsible for more than 85% of Nigeria's foreign income (Bodo & Gimah, 2019; Ebegbulem et al., 2013).

Spatial autocorrelation: A situation in which the value of a variable at a specified geographic location depends on its values at adjacent locations (Rosenberg, 2000).

Assumptions

In this study, an assumption was made that the methodology used was the best fit for the data set provided and that the data collected by the cancer registries reflected all the ethnic and minority groups within each zone. In adopting the ecological model, I assumed that the exposure variable is an aggregate of all possible environmental risk factors, including those associated with a life-course event (Clarke & Joshu, 2017). Therefore, this study assumed cumulative exposures and a life-course approach in answering the main research question. Because secondary data sets were used to answer the research questions,

it is assumed that there is no residual confounding and that the primary data collection method was consistent with the international best practices in primary data collection by cancer registries. The above assumptions meant that the outcome of this study would be accepted as true, even though it is difficult to ascertain the true effect of these assumptions on the outcome. These assumptions were necessary for this study to accommodate their potential confounding impact on the validity of the study outcome. This is because Boo and Froelicher (2013) argued that the major threats to the validity and reliability of a secondary data set arise from data collection methods of the primary data, including measurement error, sample size, and sampling frame used.

Scope and Delimitations

The scope of this study was to investigate if living in oil-producing regions is a greater risk factor in the development of colorectal cancer in Nigeria. Included in this study are all ethnic and tribal minority groups across the six zones in Nigeria. Specifically, the research problem addressed in this study is the absence of study and data on the relationship between exposure to oil pollution and CRC incidence, including studies that used quantitative, correlational ecological study design. Addressing this research problem is important to support the differential deployment of public health interventions to address the CRC disparity and motivate for positive social change in the population under investigation. The sample for this study was drawn from the populations that are covered and represented by the data set collected and managed by each regional registry and the federal ministry of health, represented by the NSCR within 2009-2013.

Only the population-based registries were included. I excluded all data from the hospital-based registries to avoid data duplication due to the migration of subjects across the zones within the data collection period. Therefore, the study was delimited to only the population covered by the population-based registries.

While the exclusion of some populations improved internal validity, it certainly challenged the external validity of the study outcome. Therefore, the outcome of this study may not be generalizable in other populations not covered by the data set used in answering the research questions. In the main, while the scope of this study covered subjects within the six regions, including all the ethnic minority populations, the study was delimited by the data, which covered only the population-based cancer registries.

Limitations

One major limitation of quantitative, correlational ecological study design is the use of regional data value or parameter to approximate an individual exposure value leading to the concept of ecological fallacy (Sedgwic, 2014). As the primary data for the NSCR secondary data set was not from a prospective study, causality is difficult because of the lack of temporal epidemiological evidence. The causality challenge can be addressed by the inclusion of ‘’Time-Trend’’ studies in future research. Also, data from non-prospective studies are vulnerable to biases (Setia, 2016). However, the use of group data and stratified ecological regression statistical analysis (Lancaster et al., 2006) may minimize the influence of ecological bias in this study. The effect of residual confounding is another limitation in this study. According to Cheng and Phillips (2014), the residual confounding in the collection of secondary data can affect the interpretation of some

variables included in the data set, affecting the validity of the data set. The absence of testing of individual potential confounding variables in ecological study design may have affected the validity of the research outcome. However, the inclusion of spatial terms like ‘’Spatial autocorrelation’’ and the adjustments for covariates using the group data in this study minimized ecological fallacy. The assumption of cumulative exposures and life-course approach in answering the main research question eliminated the potential measurement error and bias associated with the measurement at the individual levels, especially for those pollutants whose correlates cannot be easily measured by any biomarker. Therefore, most of the limitations in this study were in areas where I, as the researcher, did not have control over. In the main, the limitations of causality, biases, within-group misclassifications, the non-generalizability of the NSCR data set, and other potential confounding variables not controlled in this study may affect the overall validity of the outcome of this study, including the construct validity.

A potential barrier and challenge to this study was securing an agreement with the federal ministry of health to access the data and all the variables. As they also keep hospital-based cancer incidence from some regions without population-based data, this may confound some findings. Access to the gatekeepers to this data set was challenging and required spending some money. In addition, the ecological study design is inherently challenging as it did not allow the testing of the effect of potential confounders, especially at the individual level. Another possible limitation is the duplication of reports or data by reporting one cancer case by two different registries (Mills & Yang, 1997).

Significance

This research is significant because it tried to fill the gap in knowledge and literature and contribute to contemporary chronic disease epidemiology, including the life-course approach to disease prevention. Furthermore, the outcome of this research will inform an effective public health intervention and practice in environmental pollution management to reduce any existing colorectal cancer disparity across the six regions or geopolitical zones in Nigeria. The above is because ecological study design in this study is premised on the assumption that differences in exposure between groups are usually bigger than measurement at the individual level, and ecological studies are better to inform policy changes at the population level (Stevenson & McClure, 2005). The study outcome also has implications for social change. This is because addressing the perceived disparity in health outcomes, such as colorectal cancer resulting from differential exposure to environmental pollution, will minimize agitations and other social problems that may affect the population's health (Moses & Olaniyi, 2017). Besides, addressing any disparity in colorectal cancer across the six regions in Nigeria based on this study's evidence may support positive health behavior and other institutional reforms that will motivate social change across the country. According to de la Salonniere (2017), social change is all about the transformation of behavior, culture, social institutions, and social structure over time.

If the government appropriately implements the recommendations from this study's outcome, income from oil exports will improve as the vandalization of oil pipelines by the agitating militant youths will be substantially reduced. As mentioned earlier above, the study also promises methodological significance because of the ecological study design's

importance. The inclusion of spatial terms like ‘’spatial autocorrelation’’ will contribute new knowledge in disease distribution epidemiology. Therefore, this study will add new knowledge to the literature and inform new policies that will motivate positive social change in the population being studied. In addition, conducting spatial autocorrelation analysis will enable public health practitioners to determine if CRC incidence in the population of interest is clustered, dispersed, or occurred by chance. In the main, the outcome of this study may lead to the creation and advancement of health equity in the population under study.

Implications for Social Change

Public health interventions and social policies based on the outcome of this study may minimize the social problems of agitation, insurgences, and taking of arms, especially in the Niger Delta region of Nigeria. This is because, according to de la Salonniere (2017), social change is all about the transformation of behavior, culture, social institutions, and social structure over time. Implementing environmental pollution control and equitable distribution of resources based on differential exposures to environmental pollution will make the militants and agitating youths become positive social agents. This will eventually motivate a social change in the entire Niger Delta community, including other communities in Nigeria that are currently bearing the burden of perceived marginalization due to exposure to other forms of environmental pollution, apart from oil and its derivates. Addressing the CRC disparity in Nigeria will lead to a positive social change in the communities, especially the oil-producing communities. Any positive social change amongst the agitating militant youths will lead to the survival of the Nigerian state

(Chikwem, & Duru, 2018). In addition, addressing any disparity in colorectal cancer across the six regions in Nigeria based on this study's evidence may support positive health behavior and other institutional reforms that will reduce CRC disparity in the population.

Summary

Colorectal cancer is gastrointestinal cancer that usually starts with a polyp in the intestinal mucosa (Gramados-Romeo (2017). The annual mortality rate is about 700,000 deaths, with an estimated global incidence of 2.2 million cases in the next ten years (Bran et al., 2018). In Nigeria, like other developing countries, the incidence of 3.4 per 100,000 population (Irabor et al., 2014) is considered low compared to the average incidence rate of 17.2 per 100,000 in the European population. However, Irabor (2018), in a review of articles, concluded that there had been a steady increase in the incidence of CRC in developing countries like Nigeria. Irabor cautioned that we must accept the inevitable that CRC increase has come to stay in the West African region, where it was thought to be originally low.

Reasons for this increase are due to modifiable and non-modifiable risk factors (Ray-Ofor et al., 2020). Specifically, in the developing and low-resource countries, the risk factors are related to socioeconomic factors (Boscoe et al. 2014; Hurtado et al., 2015; Katalambula et al., 2016) and environmental-related risk factors (Bishedsari et al., 2014), such as oil pollution (Johnston et al., 2019), including the life-course events (Clarke & Joshu, 2017).

The ecological study is appropriate because differences in exposure between groups are usually bigger than measurement at the individual level, and ecological studies are

better to inform policy changes at the population level. (Stevenson & McClure, 2015). Besides, some of the exposomes associated with colorectal cancer may not have correlates, especially the complex mixtures and exposures like noise. Hence, they cannot be measured at the individual level biomarkers (Peters et al., 2012).

The study aimed to investigate if exposure to oil pollution or living near the oil-producing areas is responsible for any excess CRC outcome across the regions in Nigeria. The study also considered the possible influence of spatial autocorrelation on CRC outcomes across the six regions in Nigeria, including gender disparity. The research is of significance as the outcome will contribute new knowledge to literature in environmental epidemiology. Its outcome will also inform new policies in environmental management and differential public health interventions across the various regions in Nigeria, which has implications for social change. A detailed literature review provided further methodological, theoretical, and statistical evidence to support this study in subsequent chapters two and three, respectively.

Chapter 2: Literature Review

Introduction

CRC remains the third-most cancer globally responsible for high mortality and morbidity in developing and developed nations, with no significant gender disparity. According to Ferlay et al. (2015), CRC is the third most commonly diagnosed malignancy and about the fourth leading cancer mortality globally, accounting for over 700,000 deaths in 2012 and a projected global death of over 2 million in 2030. Researchers have reported a relatively high incidence of CRC in developed countries in comparison to developing countries. Bray et al. (2018), for instance, reported an average CRC crude incidence of about 60.30 per 100,000 in European countries, while the average CRC incidence in developing countries like Nigeria and other south Saharan African countries is estimated at 4.04 per 100,000 population (Irabor et al., 2014). Irabor et al. (2014) reported a CRC incidence in Nigeria to be 3.4 per standard population of 100,000. However, the emerging evidence indicates rising incidence cases of CRC in West African countries like Nigeria due to the increased adoption of a Western lifestyle (Irabor, 2018). South Korea, Slovakia, and Hungary were reported to have the highest incidence of CRC. In contrast, countries like Japan, Singapore, and Serbia reported the least CRC incidence in the literature review by Grandman et al. (2017).

Several pieces of extant evidence link oil pollution with poor health outcomes, such as CRC. Johnston et al. (2019) suggested a positive relationship between the impacts of oil spill exposure and various cancer sites. Chinedu and Chukwuemeka (2018) argued that heavy metals from crude oil spills are known as risk factors in poor health outcomes, while

in a landmark epidemiological study, environmental pollutants like N02 and PM 2.5 were shown to be positively associated with CRC development (Turner et al., 2017). Furthermore, long-term adverse health effects, such as CRC, have been linked to exposure to oil drilling activities (D'Andrea & Reddy, 2018). The evidence from these studies are reinforced by the research findings by Vargas (2020) that confirmed the positive association between public health issues, such as CRC, and crude oil production in the Ecuadorian oil-producing territory.

However, despite the evidence provided by several studies on the association between oil pollution and CRC, including those based on ecological studies in other populations (Lopez-Abente et al., 2012), no study so far has provided evidence on the association between developing CRC and living in an oil-producing region in Nigeria, according to my review of the literature. I also found no similar research using an ecological study approach and spatial autocorrelation analysis. Therefore, the purpose of this study was to investigate whether exposure to oil pollution by living in the oil-producing regions is responsible for any excess CRC outcome compared to living in non-oil-producing regions in Nigeria. The confounding impact of spatial autocorrelation of CRC between contiguous regions in Nigeria was also investigated. In addition, I examined the possible relationship between gender and CRC disparity as a potential measure of occupational hazards and risks within the population under study (Scarselli et al., 2018).

I begin this chapter by providing overviews of the literature search strategy and theoretical foundation. In the literature review that follows, I consider evidence related to the association between CRC and general environmental pollution, oil pollution

environment, socioeconomic status factors, lifestyle factors, gender, gene-environment interaction risk factors, and the factors related to life course exposures, including evidence on the relationship between spatial autocorrelation and CRC. A summary is provided at the end of this chapter, leading to Chapter 3.

Literature Search Strategy

I searched for literature relevant to the RQs and hypotheses. Priority was given to peer-reviewed articles. Articles not more than 5 years old were selected, except on a few occasions where earlier publications were included for cross-references. The databases of the U.S. Centers for Disease Control and a few other organization available on their websites were also searched. The articles searched focused on oil and other environmental pollution agents, including socioeconomic, lifestyle, gender, gene-environment interaction, and life-course risk factors, and their influence on CRC incidence in Nigeria and other populations. The keywords searched were *colorectal cancer and environmental pollution, oil pollution and colorectal incidence*, *cancer and environmental pollution*, *socioeconomic status and cancer incidence*, *stress and its association with colorectal cancer*, *gene-environment interaction*, *life-course epidemiology*, and *ecological studies on environmental impact on cancer health outcomes*. I searched for literature in the databases Google Scholar, SAGE Journals, CINAHL & MEDLINE, EBSCO Collection, HERO, MEDLINE with Full Text, ProQuest Health, PUBMED, as well as ProQuest Dissertations & Theses Global, at Walden University. I also used Walden's Thoreau Multi-Database Search tool. The articles I found were mainly peer-reviewed articles within 5 years of publication. The majority of the articles focused mainly on the risk factors of CRC, which

made them especially relevant in answering the three RQs. Information from relevant websites and databases that addressed the RQs are also included in the literature review.

Theoretical Foundation

I grounded this study in the ecosocial epidemiological theoretical framework. Nancy Krieger (1991) proposed the ecosocial theory, among other epidemiological theories for disease distribution. It is an emerging multilevel theory of disease distribution that integrates both social and biological reasonings. It integrates the dynamic, historical, socioenvironmental, and ecological perspectives of disease origin to inform disease distribution within a population, including the social inequalities in health. It tends to answer how socioenvironmental and biological factors interact to generate health and disease disparities within and between populations. According to Krieger (2011), ecosocial theory enabled the Faber and Harvard Cancer center working group to identify gaps in knowledge about cancer inequality by addressing the social inequality at multiple levels. I also leveraged other studies, such as those implicitly grounded in ecosocial, psychosocial, and sociopolitical theories, to inform the theoretical foundation for this study. Some of the reviewed pieces of literature anchored their theoretical argument on ecosocial and psychosocial theories, which imply that chronic disease risks are because of physical and social exposures across the life course (Ben-Shiomog & Kuh, 2002). Ben-Shiomog and Kuh's (2002) argument resonates well with the ecosocial epidemiological theoretical framework to propose an assumption that there is an integration of biological and psychosocial pathways that informs poor health outcomes over time in the population under study. Canchola et al. (2017), Kichuchi et al. (2017), and Zhang et al. (2018) all

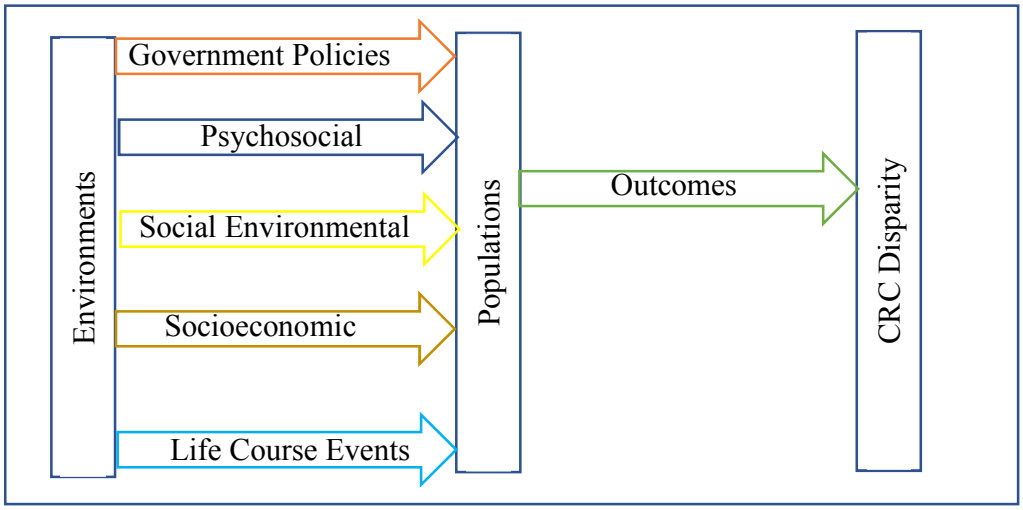
implicitly used the ecosocial theoretical framework to explain the relationship between social-environmental factors, such as living in poor built neighborhoods, stress, and low socioeconomic status, and the risk of developing CRC in some populations.

The ecosocial theory, and alternative socioepidemiologic theories of disease distribution, including sociopolitical and psychosocial theories, recognize that illness or disease is socially produced (Conrad & Parker, 2010). The three theories are premised on the fact that health and disease distribution in the human population occurs in a social context and that this distribution of disease is influenced by social, environmental, economic, political, cultural, and technological features (Krieger, 2011). Specifically, the ecosocial epidemiological theory holds that population distribution of disease, disability, death, and health and their determinants and deterrents across time and space are influenced by socioenvironmental factors (Krieger, 2001). I used the ecosocial constructs to address pathways and cumulative exposure (Krieger et al., 2018), as shown in Figure 4. I leveraged the ecosocial theory to support my independent variable as cumulative exposures beyond the individual environmental pollutants, while the dependent variable was CRC incidence cases. The preceding is because social epidemiologic theories is related to the social conditions in which people are born, live, work, and die (Marmot, 2014). The use of the ecosocial theory of disease distribution in this study also explained the reasons for a possible spatial autocorrelation of CRC between contiguous regions in Nigeria. By using an ecosocial epidemiological theoretical foundation, I was able to treat cumulative exposures in each of the six regions in Nigeria as the independent or exposure variable, including spatial autocorrelation analysis to inform CRC distribution in the populations

studied. The use of this foundation was consistent with the RQs, research problem, and the spatial terms analysis.

Figure 4

Schematic Ecosocial Theoretical Framework



Literature Review Related to Key Variables and/or Concepts

Epidemiology of Colorectal Cancer

Globally, CRC is the third most common cancer that accounts for over 700,000 deaths annually on the average, as reported in most extant pieces of literature, with an estimated incidence rate of 2.2 million cases in the next 10 years (Bran et al., 2018). In the European continent, Irabor et al. (2014) reported that the crude CRC incidence is about 60.30/100,000 population. Their male population accounts for 67.60/100,000, while the incidence in the female population is about 53.50 per 100,000 standard population. According to Irabor et al., the crude incidence rate of CRC for developing countries, especially South Saharan Africa, is 4.04 per 100,000 population. The male population

accounts for 4.38/100,000, while the female population represents a 3.69/100,000 incidence rate. In Africa, the incidence rate increases with age, which peaks at 75 years and above. Rectal cancer accounts for about 46%. These statistics show that developed countries have a higher CRC burden than developing countries.

Nigeria's incidence of CRC is largely reported as low. CRC incidence in Nigeria is about 3.4 per 100,000 population, according to Irabor et al. (2014). However, the emerging extant pieces of literature have shown that the incidence of CRC has increased in Africa, especially sub-Saharan African, like Nigeria, where the incidence was considered low by previous research reports (Rahman, 2010). Irabor (2018), in a review of articles, concluded that we must accept the inevitable that CRC increase has come to stay in the West African region. It is also instructive to note that in developing countries, where oil export is their mainstay like Nigeria, there is a large epidemiological study that linked gases associated with oil pollution such as PM 2.5 and NO2 with kidney, bladder, and colorectal cancer death (Turner et al., 2018). Moreover, other developing countries outside Africa, like the Western Asian countries, which previously and historically reported a low incidence of CRC, are beginning to report an increase in new cases and prevalence of CRC in their populations (Al-Jaberi, 1999). Besides, Basaleem and Al-Sakkaf (2004) reported the rising incidence of CRC in countries like Yemen, Jordan, and Egypt.

From the above global descriptive epidemiology of CRC, there is evidence to show that CRC incidence in developing countries like Africa, including Nigeria, is much lower than that of the developed countries (Gaylard & Ramos, 2018). The foregoing is because, as reported above, the incidence rate of CRC in Europe is 17.2/100,000 population, while

in developing countries, it is estimated to be 4.04 per 100,000 people, represented by 4.38/100,000 for the male population and 3.69/100,000 for the female population (Irabor et al., 2014). However, there is extant evidence to show an increasing rate of CRC in developing and low-resource countries, a departure from the evidence of previous studies (Colorectal Cancer Collaborators. 2019). The preceding may relate to the epidemiological transitions of CRC in developing countries (Bishehsari, 2014).

Rationale for the Selection of Variables of Interest

El-Zaemay et al. (2018) had reported that most studies on CRC were focused mainly on occupational exposures and their association with CRC incidence with not much attention on the socioenvironmental factors and life-course related factors, including the use of ecological study design. However, extant studies have linked CRC incidence to modifiable and non-modifiable risk factors (Ray-Offor et al., 2020). The major strength of using only occupational exposures is because it provides easy evidence in interventions related to immediate occupational exposures, while the weakness remains their inability to account for exposures that cannot be measured using biomarkers (Peter et al., 2012), including the life-course approach to chronic disease prevention. The majority of CRC risk factors are said to include built-environmental factors (Canchola et al., 2017), socioeconomic risk factors (Hurtado et al., 2015), environmental pollution, including pollution from oil (Kachuri et al., 2016; Johnston et al., 2019; Vargas et al., 2020), gene-environmental interaction (Valle, 2014), including life course risk factors (Chao et al., 2000; Rossi et al., 2018) and age (Gabriel et al.,2018). The above position of different researchers in this discipline formed the basis of key independent variables in this study.

In the main, the putative risk factors associated with CRC burden includes industrial environmental pollution, oil pollution, socioeconomic factors, lifestyle related factors, gender, gene-environmental interaction, and the life-course factors. Extant literature evidence related to these risk factors and CRC incidence, including the research methodologies and methods, strengths and weaknesses are discussed below:

Industrial Environmental Pollution and Colorectal Cancer

Environmental pollution from industrial sites are a risk factor in CRC incidence. Lopez-Abente et al. (2012) concluded in their research that living near a polluted industrial area increases the risk of CRC incidence. Their study aimed to investigate whether or not there was excess CRC mortality among the population living near the vicinity of industrial pollution sites in Spain. Furthermore, the socioeconomic disparity of living near a polluted environment implied the theoretical framework. The above premise is consistent with the position of Hajat et al. (2015), who argued that poor socioeconomic status is correlated with living in a polluted environment based on a global review of evidence. In the study under review, Lopez-Abente et al. (2012) used an ecological study to assess CRC mortality across 8089 Spanish municipal towns covering about 24 industrial groups between 1997 to 2006. Exposure data were obtained from the Spanish ministry of environmental and rural, and marine habitats. Poisson Regression statistical methods, like the Conditional Auto Regression Bayesian model and a mixed regression method, were used to estimate the relative risks (RR) for CRC of towns living about 2km or less and those living farther than 2km from industrial pollution sites. The results showed that living near the industrial pollution area increases CRC risk, with the highest risk coming from those living 2km or

less from an industrial mining site (RR: 1.26; 95% CI, 1.02-1.46). However, only selected pollutants were assessed in the study, excluding other potential pollutants, especially those whose correlates cannot be detected by any biomarker, such as noise and complex mixtures like crude oil (Peters et al., 2012). The concept of exposome would have improved the validity of the findings. According to Stingone et al. (2017)), the exposome is the lifelong integration of all exposures that may influence human health**.**

Two earlier studies in a Canadian population reinforced the findings of Lopez-Abente et al. (2012). However, the two Canadian studies used a small sample size to confirm a significant association between CRC and exposure to occupational agents using the case-control studies (Dumas et al., 2000; Goldberg et al., 2001). A more recent study by Oddone et al. (2014) used a meta-analysis method that included cohort and case-control studies and 83 peer-reviewed papers between 1976-2012 to confirm a varied association between CRC and the various industry exposures. Besides, Perez et al. (2016) established an association between CRC and living near the polluted industrial area, but at a 5km radius, not 2km, as reported by Lopez-Abente et al. (2012). However, the plausibility and consistency of the study by Lopez-Abente et al. (2012) are challenged by an earlier study in Sweden that confirmed no statistically significant association between CRC and occupational exposure to environmental pollution (Chow et al., 1994). Also, a more recent study in Western Australia by El-Zaemays et al., (2017) confirmed no association between CRC and 18 selected occupational agents.

Notwithstanding the inconsistencies with other study findings and the limitations associated with the use of only the Spanish population, the above findings by Lopez et al.

(2012) support the hypothesis that there is a possibility of CRC disparities within a country, based on different geographical regions' socioenvironmental factors. The study by Lopex-Abente and colleagues is of methodological significance to my dissertation as the use of ecological study can inform disparity in CRC outcomes amongst different geographical locations. Although this study by Lopez-Abente and colleagues was an exploratory ecological study, spatial terms' inclusion that minimized ecological fallacy inherent in ecological studies is a major strength. Also, the authors' recommendation that more studies are needed in other populations to improve its external validity is a gap this dissertation addressed.

Kachuri et al. (2016) conducted a study to investigate the relationship between workplace exposure to diesel and gasoline and the risk of CRC incidence in Canadian men. Included in this population-based study were 931 Colon and 840 Rectal cancer cases, including a randomly selected control from the Canadian Enhancement Cancer Surveillance System (NECSS) between 1994 and 1997. Logistic regression was used to estimate the odds ratios. The result showed that 638(36%) and 841 (46%) cases were exposed to diesel and gasoline, respectively. No significant association was recorded with colon cancer for both exposures, and gasoline exposure had no significant association with the outcome of interest. However, those exposed to diesel compared with the controls showed a significant risk ratio after adjusting for age and other potential confounding variables (OR, 1.98, 95% CI, 1.09-3.60). Kachuri et al. (2016) concluded, based on the evidence available to them, that sustained emission of diesel may be a risk factor for rectal cancer. The major limitation of this study is the use of self-reported job exposure history

that may lead to misclassification, although non-differential. An earlier study by Marie-Elsie et al. (2007) tested the association between exposure to diesel and gasoline to cancer of different sites (lungs). The findings of an association between lung cancer and exposure to diesel engine emissions reinforced the findings by Kachuri and colleagues. In this secondary analysis using data from a previous case-control study between 1979 and 1985 in Montreal, Canada, Marie-Elise et al. (2007) included 857 lung cancer cases and 533 controls, including 1,349 patients with other cancer types. All the cases and controls were all male, aged between 35-70 years. Lifetime job history was included in the model to establish any potential confounders. Notwithstanding the vulnerability of this study by Marie-Elsie and colleagues to exposure-misclassification, joined together with evidence from Kachuri et al. (2016), it is of significance to my study population in Nigeria, as there is a high population exposure to the emissions of diesel across the six regions or geopolitical zones. Since diesel is a ‘’byproduct’’ of oil refining, the above studies will support the hypothesis that living in an oil-producing environment may be a risk factor to CRC incidence in the population under study in Nigeria.

Turner et al. (2017) carried out one of the celebrated and largest epidemiological studies to assess the association between ambient air pollution and cancer mortality in the US Cancer Prevention Study II. Based on their findings, Turner et al. (2017) argued that, other than lung cancer, long-term exposure to air pollution is responsible for other site-specific cancers. The authors conducted a prospective cohort study that included 623,048 participants aged 30 and above and followed for 22 years (1982-2004). Family members of participants aged above 40 years were also included in the study protocol. Exposure to

ambient PM 2.5, N02, and Ozone pollution were assigned to the participants' residents during enrolment and follow-up. Cox Proportional Hazard models were used to estimate association after adjusting for individual and ecological covariates. After adjusting for all potential covariates, PM 2.5 pollutant was statistically significantly associated with kidney cancer (HR,1.14, 95%CI, 1.03-1.27), bladder (HR, 1.13, 95% CI, 1.03-1.23). N02 gas was statistically significantly associated with CRC (HR, 1.06, 95% CI, 1.02-1.10). The above study throws the question as to whether or not all pollutants are responsible for CRC incidence. The foregoing requires further investigation in other populations to improve the generalizability of these findings. The above study's strength is the large sample size and its long-term prospective nature, which can infer causality in the association or relationship. However, the use of cancer mortality instead of incidence as the end-point may be a limitation, as cancer mortality is only an estimation of the disease incidence and survival (Turner et., 2017).

The association between the drinking water source and CRC incidence in Jaishan county of China was tested by a study conducted by Chen et al. (2005). The authors employed a prospective cohort study. This study included 64,115 residents of Jaishan county aged 30 years and above. Participants with a reported history of cancer were excluded from the initial sample size of 75,842. All participants were divided into five cohort groups based on the sources of their drinking water. The study was initiated in 1990 and followed up till 2001. A Poisson regression statistical test was used to control for demographic and lifestyle confounding variables. Both crude and adjusted relative risks (RRs) were calculated based on the person-years follow-up period. The result showed that

‘’well-sourced’’ drinking water was statistically significantly associated with CRC, including specific colon and rectal cancer sites (RR, 1.74, 95%CI, 1.01-3.03). Based on the evidence from their research finding, Chen et al. (2005) concluded that drinking ‘’well-water’’ over a long period is a risk factor for CRC. The strength of this study remains the prospective nature of the study that can infer causality, given the temporality evidence. However, since the use of a self-reported questionnaires determined exposure status, the study is vulnerable to understating the participants’ risk. In addition, the challenge of social desirability may have also influenced the outcome. The study is of significance to my dissertation. It shows that the regions in Nigeria living in a high water-polluted environment, especially the ‘’well-water’’ source, stand a higher risk for CRC incidence.

In a more recent study, Ayuso-Alvarez et al. (2020) attempted to answer the research question posed by a social problem of exposure to industrial chemicals in the Spanish population. The authors conducted an ecological study using data from 8073 Spanish towns covering different age groups and sex. The purpose of this study was to test if there is an association between exposure to pollution from chemical plants and the risk of about 32 types of cancers using data collected between 1999 to 2008. Descriptive and inferential statistics, such as the Bayesian Hierarchical Poisson Regression model, were carried out to assess cancer risks based on a 5km radius of industrial chemical pollution. The findings suggested that living 5km or less from an organic chemical installation was positively associated with cancer incidence of the gall bladder, ovarian, and CRC, with CRC showing HR=1.09, 95%CI, 1.05-1.15. In this research findings, Ayuso-Alvarez et al. (2020) suggested that environmental exposure to some types of chemical facilities may be

responsible for the incidence of cancers, including CRC. These study findings are consistent with earlier research findings conducted by Lopez-Abente et al. (2012) in the same Spanish population. However, while Lopez-Abente et al. (2012) confirmed an association at a 2km radius, Ayuso-Alvarez et al. (2020) confirmed a similar association at a farther distance of 5km. Like most ecological studies, this study did not test association based on the individual exposures values, which may lead to ecological fallacy. The ecological study design may have conferred some strength to this study as differences in exposures in the same population are usually larger at the group level than at the individual biomarker levels (Stevenson & McClure, 2005). If living a 2km or 5km distance from an industrial pollution plant is a risk factor for CRC, it means that this study supports the hypothesis that living in the oil-polluted environment is a putative risk factor for CRC.

Living in Oil-Producing Environment and Colorectal Cancer

Several studies have used different research methods and approaches to confirm a varied association between living near or in an oil-polluted environment and the development of cancers, including colorectal cancer in different populations. According to Yuan et al. (2018), there is a positive association between living within a 10km radius of a petrochemical plant and being an elderly female and CRC risk. Based on an implied socioeconomic disparity theoretical foundation, the authors used a retrospective study conducted between 2009 and 2013 to investigate whether or not there are excess cancers for residents living in the vicinity of petrochemical plants powered by coal. The study included 2388 residential cohorts aged 35 years and above and who live within a 40km radius of a petrochemical plant. The urinary carcinogenic biomarkers of carcinogenic

metals served as exposure measures, which were retrospectively compared between those living within 40km and outside a 40km radius of a petrochemical plant. Participants were randomly selected within the Yunding county of central Taiwan. The researchers collected the relevant sociodemographic data using a face-to-face administered questionnaire survey. While the authors used both Chi-square and the student t-tests for the descriptive statistics of the discrete and continuous variables, the Poisson regression model was used to investigate the relationship between the two locations. Their findings linked cancer incidence to living near petrochemical plants. The above finding is consistent with the findings of other studies that linked living 2-5km near an industrial plant polluted area with an increased risk of CRC. This study by Yuan et al. (2018) supported the earlier occupational epidemiological studies that linked cancers of different sites with the employees of petrochemical industries (Fano et al., 2006). Yuan et al. (2018) study also supported an ecological study in the Taiwan population that suggested an association between age-adjusted cancer mortality outcomes and living near petroleum refining plants (Yang et al., 2000). This retrospective cohort study by Yuan and colleagues also supported an earlier study by Sabastian et al. (1999). The latter surmised in their research findings between 1989-1998 that high cancer incidence for those living in the Amazon Basin of Santos Carlos was linked to toxic contaminants coming from oil production activities in the Basin. Notwithstanding the relatively small sample size used by Yang and colleagues, the strength of the cohort study design and its findings addressed the methodological limitations of the previous studies that lacked the individual exposure information because of their ecological approach. The study also informed the relevance of public health

practice in environmental management, especially in the oil-producing regions, like the south-south region in Nigeria.

Nnagu et al. (2016) implied the psychosocial and ecosocial epidemiological theoretical frameworks to address the social problem of agitation and anger that is said to affect the health of the Niger Delta oil-producing communities in Nigeria. The authors used a cross-sectional study between 2013-2014 to determine the association between oil pollution and the various health problems in this population. The study included an adult population selected through purposeful sampling, aged 18 years and above, and living in the five major cities of the oil-producing communities in the Niger Delta. The outcome measures included functional capacity limitations, disease symptom inventory, environmental annoyance, and other health symptom inventories. The multiple regression analysis statistical test was used to test the relationship between all the health outcome measures of interest and oil pollution exposure. The study findings revealed that anger, worry, and intolerance were associated with oil pollution. Furthermore, the authors inferred that anger and worry could potentially trigger emotional stress, including endocrinological, immunological, and other risk factors to many health challenges, such as CRC. Nnagu et al. (2016) further inferred a cause-effect relationship between oil pollution and the constant agitation and violence in the oil Niger Delta region of Nigeria. The major strength of this study is the robust statistical analysis that included the regression diagnostic test. However, the cross-sectional nature of the study limited its ability for causal inference.

The above study by Nnagu et al. (2016) is of significance because it addressed the hypothesis in my dissertation project that is aimed at testing the relationship between living

in the oil pollution region in Nigeria and the risk of CRC. It also raised the question that supports the life-course epidemiology notion that poor health outcomes are a measure of cumulative exposures over time (Cavelin 2019). Some life-course risk factors include poverty (Boscoe et al., 2016), environmental risk factors (Bishedsari et al., 2014), neighborhood and obesogenic environment (Canchola et al., 2017), and gene-environmental interaction (Shiao et al., 2018). The study by Nnagu et al. (2016) suggested that the concept of exposome is a more practical approach to life-course studies than using the individual exposure risk factors and their biomarkers to inform the poor health-related outcomes.

Chinedu et al. (2018) conducted a secondary data analysis obtained from the Government and Non-Governmental Organizations (NGOs) websites in Nigeria to answer two main research questions related to oil pollution with a purpose to determine the quantity of oil spill in the Niger Delta region, Nigeria between 1976-2014 and also to determine the presence of toxic heavy metal exposure to the local populations. The study findings suggested the presence of heavy metals, including some toxic metals like Cadmium, Chromium, Cobalt, and Lead. Though this study did not provide a quantitative estimate of these metals, which made it difficult to determine whether or not there were within the World Health Organization set limits, the presence of these metals is of significance to the dissertation under consideration. This is because an oil spill is a major challenge in the Niger Delta region, as more than 85% of oil exploration and export in Nigeria came from this region. In addition, evidence has shown that some heavy metals, such as Cadmium, were associated with CRC for males and females in some populations

(Rogula et al., 2019). In their study to establish the relationship between CRC incidence and urban exposure to Cadmium in 18 cities in Silesia between 2010 and 2014, Rogula et al. (2019) used the R-Spearman rank test to conclude that there was a statistically significant correlation between CRC incidence and exposure to Cadmium metal.

There are pieces of evidence to link long-term adverse health effects with oil pollution clean-ups. According to D'Andres and Reddy (2018), there was a statistically significant pathological increase in the hematocrit and hemoglobin levels in participants of oil clean-ups compared with the controls (254.6+- 51.9 vs. 289 +- 63.7, $p< 0.01$). D'Andres and Reddy (2018) conducted a prospective cohort study using the exposed and the unexposed medical charts to inform this outcome. The purpose of this study was to assess the long-term adverse health effects of exposure to an oil spill following the aftermath of the 2010 Deepwater Horizon Gulf oil spill explosion in Mexico. Included in this prospective cohort study were small samples of 44 exposed subjects and 44 unexposed subjects and a follow-up of about seven years to monitor health status. Inferential t-test statistical analysis was performed to compare means between the exposed and the unexposed groups with a 0.05 predetermined significant level. Though the study was limited in the small non-randomly selected sample, the longitudinal evidence provided by this study was capable of causal inference, given the privilege of the temporality of time. The environment of the population in this study by D'Andres and Reddy (2018) is similar to the environment of the Niger Delta, and the south-south population under study as an estimated 244,000 barrels of Crude oil is spilled in the Niger Delta region every year (Ordinioha & Brisibe, 2013).

Johnston et al. (2019) conducted a review of the evidence study with the purpose to determine the impact of upstream oil extraction and environmental public health in 20 countries, including some African countries. This literature review of the evidence of multi-databases included the web of science, Ovid Global Health, EBSCO, and PubMed. The qualitative synthesis study included only 63 original peer-reviewed articles related to oil development and published between 1993 and 2017. Johnston et al. (2019) found various dimensions of associations between upstream activities and the health of the communities living near oil extraction areas compared with other populations. The mixed population in this review of evidence makes it generalizable in other populations, like the Niger Delta region in Nigeria. Even though no health outcome was mentioned in this study, it can be inferred from other previous evidence that CRC is associated with upstream oil extraction activities (D'Andres & Reddy, 2018).

Vargas et al. (2020), in a similar literature review of the evidence, found no association between oil pollution and cancer incidence, including CRC, in the Ecuadorian Amazon territories. Included in these electronic searches were PubMed, Web of Science, and Google Scholar databases, including institutional publications from the Universities and government websites. The search term inclusion criteria were mainly population studies, health effects, the petroleum industry, and the Ecuadorian territories. The finding from this study contradicted other previous studies, such as those conducted with a similar review of the evidence (Johnston et al.., 2019) and prospective cohort study (D'Andres & Reddy, 2018). I hold the premise that the evidence provided by Vargas (2020) should be interpreted in the light of superior evidence provided earlier by D'Andres and Reddy

(2018) through a more causal prospective study design and Johnston et al. (2019) that included different populations. Both D'Andres and Reddy (2018) and Johnston et al. (2019) used a more varied population that guaranteed external validity when compared to the work by Vargas et al. (2020). In the main, living in an oil-polluted environment is a risk factor for CRC and other poor health challenges. The above position is reinforced by a more recent study that confirmed the positive association between oil spills and gas flaring with adverse pregnancy outcomes, such as hypertension (Oghenetaga et al., 2020). The above findings by Oghenetaga et al. (2020) provided some preliminary evidence supporting my dissertation study's the hypothesis in that living in an oil-producing environment may be a putative risk factor for CRC incidence in the Niger Delta region of Nigeria. The above findings may also be of significance to support the upward review of the present financial compensation for those living in Nigeria's oil-polluted environment, as suggested by a recent scholarly research submission by Nnaemeka (2020).

Ray-Offor et al. (2020) conducted a study in the Niger Delta populations to investigate the frequency and presentations pattern of CRC in two oil-producing areas of Rivers and Bayelsa Niger Delta states, Nigeria. The observational study was carried out between 2013-2018 in public health facilities in the population studied. Data were collected through questionnaires from practicing endoscopists. The result confirmed 44 CRC cases, with the male population representing 66% in a study population of mean age of 52.3 years. The authors concluded that CRC is not uncommon in the two Niger Delta oil-producing states. The study outcome of Ray-Offor et al. (2020) confirms the presence of CRC in the oil-producing region. However, this study was not a correlational study and, therefore, did

not provide any aetiology as to CRC incidence , which is useful in my study. In the non-oil-producing communities. Irabor (2014) conducted a study in Ibadan, a non-oil producing southwestern state in Nigeria. The study included a retrospective study that extracted data from studies carried out between 1995-2004. The outcome of this research showed the increasing prevalence of CRC in non-oil-producing communities. Irabor (2014) study findings were reinforced by a similar study by Ibrahim et al. (2013), which used the same review of evidence to confirm the increasing cases of CRC in non-oil-producing populations of Ilorin, Kwara State, Nigeria. The study also confirmed the rising cases of CRC amongst younger populations aged below 40 years. Put together, the above-reviewed studies revealed the rising cases of CRC in both oil-producing and non-oil-producing communities in Nigeria. However, none of the studies in Nigeria used a correlational method to establish relationships or causal inference between social-environmental factors and CRC outcomes at the various regions in Nigeria under study. The foregoing is the gap that this dissertation will aim to address to establish putative risk factors that are responsible for any disparity in CRC outcomes in Nigeria.

Particulate Matter (PM 2.5) Pollutant and Colorectal Cancer

Particulate matter (PM 2.5) is one of the environmental pollutants responsible for cancer of different sites, including CRC. The relatively small particle size and its ubiquitous nature make it a risk factor for different cancer types. Su et al. (2019) conducted a study to establish the geographical distribution and the relationship between air pollution and cancer incidence in the Taiwan population. The study was an integrated ecological study design covering 19 cities and 349 local administrative areas (LAA). This study

included cancer cases obtained between 2012 and 2016, extracted from the Taiwan Cancer Registry. Air pollution data were obtained from the Environmental Protection Administration of Taiwan. Statistical analysis involved the Spearman rank correlation coefficient in estimating the level of correlation between the age-adjusted CRC incidence rates and the various air pollutants, such as PM 2.5 after stratification into gender and urbanization degrees. The findings showed that PM 2.5 was associated with various cancer incidences in the Taiwan population, with higher associations recorded in southwestern and mid-western Taiwan. The above study suggested that, indeed, PM 2.5 pollution was responsible for different types of cancer incidence. The above finding informed PM 2.5 as a putative risk factor for CRC in the dissertation project. More recent studies have narrowed down the association between PM 2.5 and cancer to CRC.

Jemwitheesuk et al. (2020) conducted a retrospective cohort study to specifically determine if there is any statistically significant spatiotemporal association between air pollution exposure and the risk of CRC in the Thailand population. CRC cases were extracted from the cancer database collected between 2010 and 2017 and kept with the ministry of health, while the air pollution data collected between 2010 and 2016 were obtained from NASA's air pollution exposure databases. The entire NASA database was compiled from the database of the aerosol diagnostic model. Included in this study model was 59, 605 colon cancer patients from the various provinces of Thailand. The Bayesian hierarchical spatiotemporal and the poison log-linear statistical models were used to estimate the incidence rates of CRC. The result showed that the posterior probabilities of risk appeared the most in the dust-PM 2.5, representing about 15%, while the black and

organic carbon pollutants accounted for 4% each for every increase of 10ug/m3 of each pollutant candidate. The result from the statistical analysis of the data from this study suggested that PM 2.5 was associated with CRC incidence for every 10ug/m3 increase. The finding from Jenwitheesuk and colleagues reinforced the earlier study by Su et al. (2019) that demonstrated a generic association between PM 2.5 and various cancer types. The two studies above support the research hypothesis of the main research question under investigation in this dissertation.

To provide further temporal evidence for causality, Chu et al. (2020) conducted a prospective study to determine the association between PM 2.5 and CRC and its sub-types in the United States population. 139, 534 cancer-free patients from ten United States research centers were included in the study. Participants were followed up for ten years. The Cox regression statistical model was used to estimate the risk of PM 2.5 exposure and CRC incidence and its sub-types. After adjusting for all the potential confounders, the study found that a 5ug/m3 increase in PM 2.5 was statistically significantly associated with CRC (HR=1.27, 95% CI: 1.17-1.37). The genome-wide interaction analysis showed that a multiplicative interaction occurred between PM 2.5 and genetic variants of CRC. The strength of this study is the temporal evidence and the large sample size. Using a single pollutant model in this study is considered a strength, as it provided further evidence to support PM 2.5 as a single putative risk factor for CRC. Therefore, long-term exposure to PM 2.5 may lead to CRC incidence, which genetic variants might modify. Apart from the temporal evidence and the use of a single pollutant in this study, the use of long-term exposure also supports causality and the assumption of cumulative exposures in the

dissertation's research design. The above study by Chu et al. (2020) is also supported by another time-series study by Ethan et al. (2020). Ethan and colleagues used a time-series study design to investigate the possible association between exposure to Pm 2.5 and mortality due to stomach cancer and CRC in the Xian population. The study's findings revealed that PM 2.5, as a single pollutant, was associated with stomach cancer mortality but when in combination with NO2 pollutant, was statistically significantly associated with CRC mortality (RR=1.015, 95% CI: 1.009-1.017).

Guo et al. (2020) conducted another cohort study with a larger sample size than Chu and colleagues. In this large cohort study in the Taiwanese population between 2001to 2014 and followed-up until 2016, Guo and colleagues included 385 650 Taiwanese adults aged above 18. Exposure data were obtained at the individual level based on a 1 km resolution spatiotemporal model. The outcome variable of interest (CRC) was obtained from the national death registry in Taiwan. The Cox-proportional hazard regression statistical model was used to estimate the association between PM 2.5 and deaths from gastrointestinal, CRC, stomach, and liver cancer sites. The results showed that for an increase in 10ug/m3, PM 2.5 was marginally associated with CRC after adjusting for the included covariates (HR=1.13,95% CI: 1.00-1.36). One of the main limitations of this study is that the PM 2.5 exposures were calculated on the basis of residential pollution, which excluded exposures at the workplaces. Hence, the total estimate of risk may have been understated. However, the strengths remain the large sample size, prospective nature, and long-term follow-up, which can confer causality between the variables under review.

The limitation may be overcome in this dissertation because exposures are measured at the regional levels that will include pollution in the residential and workplace environments.

Overall, only a few studies, such as Wong et al. (2016), reported no association between PM 2.5 and CRC using a cohort study in Hong Kong. Also, Turner et al. (2017) reported a positive association only with a combination of PM 2.5 and NO2 exposures and CRC incidence and not with PM 2.5, as a single pollutant in one of the largest cohort studies in the United States. However, most of the studies, which were mainly prospective cohort studies, reported an association between PM 2.5 and CRC in different populations. The evidence of temporality, consistency, and external validity provided by these studies that showed a positive association between PM 2.5 and CRC incidence, supports the research hypotheses in my dissertation that living in an oil-polluted region could be a single putative risk factor for CRC incidence in Nigeria, especially those living in the oil-producing regions. The above is further supported by studies that link PM 2.5 pollution to the oil and gas industry (Thompson et al., 2016; Fann et al., 2018; Viette et al., 2020), like the population under study in this dissertation. Therefore, PM 2.5 is considered a putative risk factor for CRC, and its inclusion as exposure or predictor or independent variable in this study is justified.

Gender and Colorectal Cancer

There are pieces of extant evidence to suggest that gender may be a risk factor in CRC incidence across populations, including disparities in age, race, and cancer sites. Gao et al. (2008) conducted a retrospective cohort study in the Canadian population between 1981 and 2001 to establish gender differences in CRC incidence, mortality, hospitalization,

and surgical procedures in Canada. In this study, each CRC outcome was identified using the international classification of diseases (ICD). Included in this study was an age-standardized CRC rate using the 1991 Canadian population census. Ms-Excel was used to calculate the moving average. The study findings suggested that the overall CRC incidence was higher in men than women. While the male had higher incidence rates for rectal cancer, the female had a higher incidence for the right-colon sub-types. The findings suggest that a CRC screening intervention that reflects gender disparity is imperative. This study is also significant in the dissertation as it supports the third research question that examined gender disparity in CRC outcomes. The study's outcome under review may also explain the possible occupational risks in the oil-producing region in Nigeria, where men are more involved in oil exploration than women because of the occupational hazards associated with offshore oil exploration.

Nguyen et al. (2009) conducted a systematic review of evidence and meta-analysis to provide quantitative pooled risk estimates of the association between gender and advanced colorectal neoplasia. Included in this study were 924 932 men and women from 18 different populations. This study also included subjects with average risk and asymptomatic individuals undergoing screening colonoscopy and subjects with a family history of colorectal neoplasia. The study's primary outcome measure was the relative risks of advanced neoplasia, while the secondary end-point was the relative risks for CRC. The results showed a statistically significant association between gender and advanced CRC neoplasia from age 40 years and above 70 years (RR=1.83; 95% CI: 1.69-1.97). The meta-analysis study concluded that men were at more significant risks for advanced colorectal

neoplasia than women across all age groups. This finding is important in informing gender-specific CRC screening interventions across populations. The large sample included in this meta-analysis was the major strength of this study. This study under review also supports the earlier study in Canada by Gao et al. (2008) that found a positive association between gender and CRC incidence.

In a similar review of evidence study by White et al. (2018), the authors used a cross-sectional review of national data on CRC incidence rates by age obtained from 2010 and 2012 in the UK to conclude that men had a higher risk of developing CRC than the women. The study findings by White et al. (2018) is consistent with the previous study by Yang et al. (2017) that used a meta-analysis study of 13 retrospective cohort studies and one randomized controlled trial to demonstrate that sex of a person was a single-significant predictor of CRC survival in different populations. Also, in an earlier large retrospective cohort study in the United States, Abotchie et al. (2011) included 373 956 CRC patients aged above 40 years to confirm that CRC incidence was higher in men than in women. Even though this study showed no significant difference in the crude CRC incidence between men and women, men showed a higher CRC incidence rate across age, race, site, and CRC stage. The result also revealed that CRC gender disparity was increasing with age. However, the gender disparity with age narrowed down over time. Also important in this study is the evidence that older women more than 65 years old were at a higher risk than men in the same age cohort. Women have also been shown to present with right-sided colon cancer than men (Kim et al., 2015).

In the main, the evidence reviewed in the literature suggested that gender is a risk factor in CRC development. Therefore, CRC could also be associated with occupational risk factors. The finding of gender disparity is important in informing differential CRC screening interventions, including allocating resources to accommodate gender differences.

Socioeconomic Status and Colorectal Cancer

The neighborhood environments where people live have been implicated as a risk factor in cancer development, including the CRC. Kim et al. (2010) premised their research study on the ecosocial and psychosocial theoretical foundation to provide evidence of a positive association between socioeconomic status and the risk of colon and rectal cancer in women. Kim et al. (2010) used data from a prospective cohort study that included 111, 129 women aged 30 to 55 years in the ‘’Nurses’ Health Study’’ between 1986-2006 using the US-Census- derived characteristics of a black group of residents. Cox proportional hazard model was used to estimate the association between neighborhood socioeconomic status (SES) and the incident colon and rectal cancers, including the effect modification. The individual risk factors were included as mediators in the relationship. This study found no association between the neighborhood SES and colon cancer for all women. However, higher SES was inversely associated with colon cancer. The result also revealed that education interacted with neighborhood SES and BMI while the intake of red meat mediated in the relationships above. The researchers concluded from their research findings that living in a higher SES environment may protect against rectal cancer in all women and colon cancer in only women with higher education, mediated by some

behavioral risk factors. The strength of this study is the prospective nature of the secondary data set and the large sample size. The limitation of this study is its vulnerability to exposure misclassification, which may be largely non-differential. Future research is recommended to focus on estimating discrepancies in the behavioral risk factor epidemiology for both colon and rectal cancer. The above study findings support the inclusion of covariates analysis in this dissertation project.

The evidence provided by Kim et al. (2010) was reinforced by several studies that confirmed a positive association between living near a poor neighborhood defined by its proximity to industrial pollution and the risk of cancers, such as CRC (Lopez-Abente et al., 2012; Ayuso et al.,2020). Boscoe et al. (2016) also provided evidence to link living in a low SES environment related to poverty, and the incidence of CRC in the low-income population. The question is: If the low SES environment is positively associated with CRC, why is the prevalence still higher in those living with high SES neighborhoods? I argue that other intervening factors, such as genetics, gene-environment interaction, lifestyle, etcetera, may be responsible for the risk or otherwise protective effects to account for this disparity within the socioeconomic divide. The previous is instructive and motivates ecological studies and a life-course epidemiology approach that accounts for all exposures to inform CRC and other chronic disease outcomes.

Freedman et al. (2011) conducted a study to determine an association between the neighborhood and chronic disease onset in adult life in the US population. Freedman and colleagues used the 2002 wave, which reflected the year 2000 US Census boundaries. In this study, neighborhood factors, such as environmental, social, crime, segregation, built

environment, and health care delivery were modeled for adults aged 55 years and above, who survived the 2004 wave using the retirement study scales, which was validated for built, social, and economic environments. The outcome variables included in this study were cancers, heart, and diabetes diseases. The hierarchical logistic regression model was used to estimate the pseudo-intraclass correlation coefficient (ICC) to determine the variability in outcome attributable to neighborhood variations after controlling for the individual and neighborhood-level factors. The results suggested that a living neighborhood prone to a crime was positively correlated with cancer development (OR=1.18, 95%CI, 1.03-1.37). Also, living in an economically advantaged area was positively associated with lower odds of developing the onset of diabetes (OR=0.80, 95%CI, 0.69-0.94). The study concluded that the neighborhood environment might determine the incidence of cancer, such as CRC. This is because crime-prone neighborhoods might lead to high stress and the associated cortisol levels, which can mediate chronic diseases like cancer, including CRC. This preceding research supports the positive association between the neighborhood and chronic disease onset at adulthood.

The research finding by Freedman et al. (2011) was reinforced by Kruk et al. (2019) confirming the positive relationship between psychological stress and chronic diseases like cancer. Furthermore, there are other pieces of evidence to support a positive association of CRC with the built environment (Canchola et al., 2017; Doubeni et al., 2012), perceived stress (Kikuchi et al. (2017), and other environmental risk factors (Bishedsari et al. (2014). In conclusion, the study findings by Freedman et al. (2011) supported the hypothesis in my dissertation that adulthood chronic diseases onset, like CRC, may be a function of the

cumulative effect of different exposures in a given geographical environment or population across the life course.

Doubeni et al. (2013) conducted a study using data from an ongoing large prospective cohort study from the National Institute of Health-AARP Diet and Health Study. This study aimed to determine the relationship between the individual and area-level SES and CRC by location and overall. 506, 488 participants aged 50-71 years and living in the United States between 1995-1996 were included in the study. While the information on the outcome variables was obtained from the Cancer Registry, the data on SES was obtained through a self-reported questionnaire. The result from this study indicated that out of the total 506, 488 participants, 7676 were found to have invasive CRC (44.6% in the right colon, 26.7% in the left colon, while 25.5% were found in the rectum. CRC incidence was also found to be higher in those with low education and living in the low-SES neighborhood. Based on the results, it was concluded that SES was significantly associated with CRC risk (Doubeni et al., 2013). The prospective cohort study under review had the potential to infer causality, while the main limitation of this study is the use of a self-reported questionnaire for the SES exposure. This is because self-reported data is vulnerable to both self-reported and social desirability bias (Althubaiti, 2016). The findings support the assumption in my dissertation project that the six different regions or geopolitical zones in Nigeria with different socioeconomic and ecological assets may have an inherent disparity in their CRC risks. The finding of this study is supported by the previous study by Freedman et al. (2011) who used the data from those survived the 2014 wave in the same United State population, and aged 55 years and above, to conclude a

positive association between neighborhood-related exposure factors and chronic disease onset in later life. Notwithstanding the influence of social desirability and self-reported biases, this study's finding is significant to public health practice. It will provide the opportunity to correct CRC disparities by addressing the SES risk factors as neighborhood environmental risk factors.

Canchola et al. (2017) conducted a study to investigate if there is an association between obesogenic environment attributes and the risk of CRC. The study was a secondary analysis of data set containing 81, 197 participants (35,397 male and 45,800 female) living in California, USA, identified between 1993-2000. Cox Proportional Hazard models were used to estimate hazard ratios for the overall CRC risks and the variations in risk among racial/ethnic groups in the population. The study found an association only between CRC and high traffic density environment in males (HR=1.25, 95%CI: 1.02-1.61, p=0.03). The study did not confirm any disparity in CRC outcome among ethnic/racial groups within the context of the environmental risk. The large sample size supported the evidence in this research, but, like most secondary analyses, the validity is challenged by the potential for type 2 error (Clarke & Cossette, 2000). This study supported the ecological study in Nigeria population, given the high traffic nature and the associated health outcome across all regions in Nigeria.

Kikuchi et al. (2017) confirmed an association between perceived stress and the development of CRC. Their prospective cohort study included 110,585 Japanese adults (46,395 men and 64,190 women) aged between 40 and 79 years. It was conducted in 45 locations in Japan between 1988 and 1990. Cox Proportional Hazard models were used to

estimate hazard ratios for colon and rectal cancers. The study findings suggested a positive association between perceived stress intensity with rectal cancer (HR: 2.94, 95%CI: 1.09-7.93) only in men and perceived stress frequency as a protective factor for colon cancer (HR, 0.23, 95%CI: 0.07-0.73) only in women. No statistically significant association was found between perceived stress frequency for both colon and rectal cancer in men. The study also found a positive association between perceived stress and CRC, but the association did not show any consistent pattern. The above statement poses the challenge as to the use of CRC as an outcome measure instead of using the colon and rectal outcome measures. According to Kikuchi and colleagues, using CRC as an outcome measure will not allow for a distinction between risk and prognostic factors (Kikuch et al., 2017). Hence, the inconsistencies in the pattern of the positive association with CRC. However, I hold the premise that the use of rectal and colon cancer as separate outcome measures may lead to mutual misclassification (Chao et al., 2000). The findings of this study are consistent with the premise in this dissertation that those living in the oil-producing region in Nigeria constantly complained of psychological and perceived stress due to the perceived marginalization and the lack of adequate compensation for providing more than 85% of the federal government revenue accruing from crude oil export (Moses & Olaniyi, 2017).

Orlewska et al. (2018) used a cross-sectional ecological study to assess the association between frequent cancer sites and socioeconomic variables in Poland. Cancer incidence was calculated on the basis of the national Polish Cancer Registry. The individual SES was measured using the 2015 Polish social cohesion survey. The Spearman Correlation Coefficient was used to test the association between the incidence of cancer

and SES factors at 0.05, 2-tailed significant level. The findings suggested that friends and neighborhood social capital were statistically significantly negatively associated with colon cancer, while poverty was statistically significantly associated with lung cancer. The findings of this study should be interpreted with caution or within the context of the limitation of a cross-sectional ecological study design. However, the finding that colon cancer was significantly associated with neighborhood social capital resonates well with any proposed public health intervention that will increase the ecological asset and social capital base of all the communities across the regions in Nigeria. The ecological study design is good in providing evidence for public health policies that will address any disparity in health outcomes in each population (Stevenson & McClure, 2005).

Cigarette Smoking, Alcohol Consumption, Fruits and Vegetables Intake, Physical Activity, and Colorectal Cancer Risks

Cigarette smoking is one of the lifestyle factors that has been linked with CRC outcomes by many research studies. According to Chao et al. (2000), long-term cigarette smoking is associated with the increased risks of CRC mortality in men and women. The researchers used data from a nationwide prospective cohort study of the cancer prevention study initiated by the American Cancer Society (ACS), which included 312,332 men and 469,019 women recruited across the 50 states of the United States and some US territories. The cohort age ranged from 30 years and above, with a median age of 57 years for men and 56 years for women. Follow-up was between 1982-1996. The Multivariate Cox proportional hazard statistics model was used to estimate risks. Race, BMI, educational level, family history, use of aspirin, and vegetable intake, were adjusted for as covariates

or confounding variables. After follow-up, the results revealed a study end-point of 4432 colon and rectal cancers. The multivariate-adjusted CRC mortality rate was highest among current smokers compared to never-smokers (RR=1.33, 95%CI: 1.16-1.45) for men, and (RR=1.41, 95%CI: 1.36-1.58) for women. The increased risk was higher after 20 years or more for both men and women. What is instructive in this study is that the researchers selected the covariates on the basis of their previously observed association with CRC mortality and cigarette smoking and based on the previously reported risk factors associated with colon cancer mortality in Cancer Prevention Study II (CPS). I surmise that the result of this study can approximate causal inference because of the prospective nature and the large sample size. However, self-reported smoking exposure may understate smoking risks because of social desirability (Althubaiti, 2016). Smoking as a coping mechanism may be a mediating factor in the CRC in communities with perceived stress due to constant agitation, as the Niger Delta region in Nigeria. According to Siopen et al. (2013), smoking is associated with psychological stress and perceived inequality, leading to poor health outcomes. The above evidence is important for public health practice in providing interventions that will reduce perceived stress and inequality in the Niger Delta population in Nigeria, which may reduce any CRC disparity over time.

In a related study, Hannan et al. (2009) used the same Cancer Prevention Study, otherwise known as the National Cohort, like the previous study by Chao et al. (2000), to assess the association between cigarette smoking and the risk of CRC in the United States. While Chao et al. (2000) focused on the relationship between cigarette smoking and CRC mortality, Hannan et al. (2009) focused on the risk of CRC incidence in mainly the White

population. Included in the study by Hannan and colleagues were 86,402 men and 97, 785 women. The CRC outcome was identified by the international classification of diseases (ICD-9 codes 153 -154.1 or ICD-10 codes C18-C20. 9). After adjusting for all known and putative risk factors, the Cox proportional hazard analysis showed that CRC incidence was statistically significantly higher in current smokers (HR=1.27, 95%CI: 1.06-1.52) and former smokers (HR=1.23, 95% CI: 1.11-1.36), when compared to the lifelong non-smokers. Causality was possible in this study because of the large sample size and prospective cohort nature. However, the US Surgeon General ruled out any causality, which may be related to the threat of residual confounding and the use of self-self-reported exposure variables. The use of mainly the White population would have also challenged its external validity, even within the US population. The consistencies in the findings of Chao et al. (2000) and Hannan et al. (2009) using temporal evidence and reinforced by the findings of Wei et al. (2011), who confirmed that 20% of CRC were attributable to cigarette smoking, improved the causal inference of this association in line with the Hill's criteria for causation. However, the population used in this study may not allow the findings to be inferred in the Nigerian population, which is mainly African-blacks.

More recent prospective studies, including those with quantitative estimation, confirmed cigarettes smoking as a risk factor in CRC incidence, including site-specific incidences. Hurley et al. (2013) used large data from a prospective cohort study that included 122,264 female participants in California and followed up from 1995-2009. The 1,205 cases (917 colons and 288 rectal), prospectively diagnosed, were identified through the California Cancer Registry. Exposures included only participants who have actively

smoked 100 cigarettes or more. Hurley et al. (2013) concluded that long-term smoking was a risk factor for CRC based on the 30% increase in the relative risks compared to non-smokers. In a different population, Parajuli et al. (2014) also used data from a Norwegian prospective cohort study that included 602, 242 participants (49% men and 51% women) aged between 19-67 years, who had earlier enrolled in the previous Norwegian Health Survey between 1972-2003. After 8.6 million person-years of follow up, Parajuli et al. (2014) confirmed that the ever smokers had a statistically significant increase in rectal cancer in men (HR: 1.27, 95%CI: 1.11-1,45) and (HR: 1.28, 95%CI: 1.11-1.48) in women, compared with gender-specific never smokers. Parajuli also confirmed in this study that the number and frequency of cigarette smoking were positively associated with rectal cancer, with no heterogenicity by gender. In another prospective cohort study conducted across ten west European countries that included 521,330 men and women aged between 25-70 years of age, Alek-Sandrova et al. (2014) concluded that adherence to five lifestyle risk-reduction behaviors was associated with the reduction of CRC. The dichotomization of the lifestyle variables may have led to methodological challenges and loss of power in this research. However, using different European populations improved its external validity, though not in African countries, like Nigeria. In a similar research topic that involves the investigation between lifestyle factors and the risk of CRC at index colonoscopy in a ‘’Fecal Immunochemical Tests (FIT) in Bologna, Italy, Colossi et al. (2018) included data from 3894 FIT positive patients aged 50-69 years, and collected between 2005 and December 2013 to answer their research question. The findings showed that smoking as a lifestyle factor was independently associated with CRC (OR=1.50,

95%CI: 1.10-2.04). Alcohol consumption and obesity were also independent risk factors (OR=2.29, 95%CI: 1.15-4.58) and (OR=1.29; 95%CI: 1.05-1.60), respectively. Colossi and colleagues concluded that various lifestyle factors are associated with the risk of CRC. Notwithstanding that no control group was included in the model, this study was still one of the leading pieces of evidence in the causal pathway between CRC and the various modifiable risk factors. Put together, all the studies have provided evidence to support a positive association between CRC and modifiable risk factors like smoking, which are biologically plausible, consistent, and temporality evidence to approximate causality. The preceding provided evidence that cigarette smoking is a risk factor for CRC. Alcohol is also considered another modifiable risk factor, which is discussed below.

Alcohol consumption has been linked to having a direct effect or through interactions with other lifestyle factors with CRC development. According to Rossi et al. (2018), epidemiological studies show that alcohol consumption is a risk factor for CRC incidence through the generation of acetaldehydes and other ac0hol metabolites that lead to the activation of cancer-promoting cascades, including the DNA-adduct formation, oxidative stress, peroxidation of the lipid, and alterations in the epithelial barriers and their inherent immune-modulating effect. Colon cancer associated with alcohol consumption is assumed to be due to low-intake of foliates, poor diet, and circadian disruption, which induce colon carcinogenesis when combined with other factors (Dinis-Olivera, 2016).

Pedersen et al. (2003) conducted a cohort study to determine the association between the amount and type of alcohol and the risk of colon cancer in the Danish population. The researchers included randomly selected 15,491 men and 13641 women

aged 23-95 years old. Incident cancer information was collected from the Danish cancer registries. Exposures included beer, wine, and spirits, which were obtained through a self-completed questionnaire. Pedersen et al. (2003) used the Cox proportional hazard model to estimate the effect of alcohol on CRC, with age as the underlying time scale. Non-drinkers were entered into the model as the reference group. Within the 426,934 person-years, 411 colons and 202 rectal cancers were prospectively diagnosed. No interaction occurred, among other risk factors, with alcohol. The findings suggested a positive association between alcohol consumption and the risk of rectal cancer with no statistically significant association with colon cancer. However, the inclusion of wine reduced the carcinogenic effect of alcohol consumption. The use of a population-based prospective cohort study with a large sample size was the major strength in this study, notwithstanding the threat of social desirability effect, occasioned by the use of a self-completed questionnaire on the exposure factors. Jim et al. (2013) identified both risk and protective effects of alcohol consumption, which were dose-dependent. In this methodological review study, Jim et al. (2013) conducted sensitivity and cumulative meta-analysis that included18 cohort studies. The researchers surmised in their finding that heavy alcohol drinking up to 50g/day was positively associated with CRC (RR= 1.31, 95% CI: 1.23-1.39), while light drinking had a protective association (RR=0.91, 95% CI: 0.89-0.94). The studies by Pedersen et al. (2003) and Jim et al. (2013) confirmed that alcohol consumption was a risk factor in CRC, especially at a high dose.

Extant literature has also linked life-long alcohol consumption with genetic mutation and pathological sub-types and pathways to inform CRC incidence. Dashti et al.

(2017) confirmed that more than 28g/day alcohol consumption was associated with CRC risk in the DNA-Mismatch Repair (MMR) gene carrier population. In a study that involved the use of the US colon cancer family registries obtained in the US, Canada, and Australia, Dashti et al. (2017) included 1925 subjects with MMR gene mutation between 1997-2012. 769 CRC cases were diagnosed at a mean age of 42.6 years. With age as the time scale, Cox proportional hazard analysis showed that the consumption of 28g/day of alcohol was associated with an increase in CRC risk (HR=1.69, 95% CI: 1.07-2.65) when compared with non-alcohol drinkers. Rectal cancer showed no association, while colon cancer was statistically significantly associated with alcohol (HR=1.94, 95% CI: 1.19-3.18). This study among the MMR gene mutation carrier population is significant. However, a large prospective cohort study is needed to inform any causality and improve internal validity. Zaitsu et al. (2019) confirmed that long-term alcohol consumption was a risk factor for CRC. The researchers conducted a nationwide, multicenter, hospital-based case-control study in Japan between 2005 and 2016, which included 63,232 cases and 63,232 hospital controls, aged between 20 years and above. The study also suggested a dose-response relationship between alcohol consumption and CRC compared to lifetime abstainers (Zaitsu et al., 2019). Even though case-control studies are vulnerable to exposure misclassifications, this large population-based case-control study is significant to inform evident-based public health practice to reduce chronic abuse of alcohol in the Japanese population.

In a more recent case-control study in the Darmkrebs, Chancen der Verhutung Durch (DACHS) study initiated in 2003, Amitan et al. (2020) included 2444 cases and

2475 controls. In this study, the tumor was analyzed for Microsatellite Instability (MSI) and Kristen Rat Sarcoma (KRAS). The multinomial statistical analysis showed that alcohol consumption was a risk factor for MSI-high and KRAS-high related CRC outcomes, though smoking showed a stronger association. Alcohol consumption is common in Nigeria, especially among those in the Niger Delta region, which is argued to be associated with perceived stress related to environmental oil pollution and marginalization. According to Tavolacci et al. (2013), perceived stress is positively correlated with alcohol consumption as a coping strategy. The above evidence is significant in informing the use of cumulative exposures as the independent variable in the estimation of the relationship between living in the oil-polluted environment and the risk of CRC. Alcohol consumption is indeed a modifiable risk factor for CRC. Another modifiable risk factor, like physical activity, is discussed below.

Evidence has shown that physical activity, including intense activity and frequency, confers some protection against CRC. In the research report, Slattery et al. (2004) argued that intense activity is a protective factor for CRC by increasing gut motility, enhancing the immune system, decreasing insulin growth factor levels, reducing obesity, and removing free-radicals. Thune and Furberg (2001) conducted a study to determine the relationship between physical activity and the overall site-specific cancer risks using the observational study design that included both the cohort and case-control studies. This study showed that both leisure-time and occupational-physical activity were protective of the overall cancer risks, with dose-response inverse relationships for both sexes. A randomized clinical trial would have been included to determine the possibility of causal

relationships between physical activity and its role in site-specific cancer outcomes. Slattery et al. (2003) used a cross-sectional study between 1977-2002, which included 952 rectal cancer cases and 1205-age and sex-matched controls. Multiple logistic regression statistical analysis showed that, in both men and women, intense physical activity was a protective factor for rectal cancer (OR=0.60, 95% CI: 0.44-0.81) for men and (OR=0.59, 95% CI: 0.40-0.86) for women. The findings of this study are of significance in the dissertation study because, in some regions, like the Niger Delta region in Nigeria, their claimed loss of land to oil pollution and degradation may result in the lack of built environment for physical activities and walkability. The above may predispose them to a higher risk of CRC, given the protective role of physical activity in CRC (Slattery et al., 2003; Thune & Furberg, 2001). The above findings are consistent with the World Cancer Research Fund (2018) position that used global scientific and research experts through a review of evidence to inform the protective effect of physical activity on CRC incidence globally. Physical activity plays a protective role against CRC. The protective role of the consumption of fruits and vegetables is discussed below.

Park et al. (2007) conducted a study to determine any association between fruits and vegetable intake and CRC risks in the US population. The authors used the National NIH-AARP Diet and Health Study conducted between 1995-1996 and included 488,043 men and women aged between 50-71years. A self-reported questionnaire collected exposure data on fruit and vegetable consumption. After a follow-up of 2,121,664 person-years, 2972 incident cases (2048) for men and (924) for women were diagnosed. Park et al. (2007) concluded that the intake of fruits was not correlated with CRC for both men and

women. However, Vogtmann et al. (2013) confirmed that the intake of fruits was a protective factor for CRC, while no statistically significant association was found for vegetable intake. The above study conducted in Shanghai in China included 61, 274 males aged 40-74 years. The exposure variables included eight fruits and 38 vegetables. After a total follow-up period of 390,688 person-years, 398 cases of CRC (236 colons, 162 rectal) were diagnosed. Cox proportional hazard analysis revealed that fruit intake was a protective factor for CRC (HR=0.67, 95% CI: 0.48-0.95), while vegetable consumption did not show any significant association. Kashimo et al. (2015) used a systematic review and meta-analysis of previous case-control and cohort studies to determine specifically the relationship between vegetable intake and CRC risks. Based on the epidemiological evidence and biological plausibility, Kashimo and colleagues did not reach a conclusive inference. The above findings of no statistically significant evidence to support the association between vegetable intake and the risk of CRC were reinforced by Lin et al. (2005), who found no association between vegetable intake and CRC in a prospective cohort study in the US population that included 39, 876 healthy women aged above 45 years. Also, in another prospective cohort study in Japan, Sato et al. (2007) found no association between vegetable and fruit intake and CRC risks. However, Kunzmann et al. (2015) found a rather weak and inconsistent association with vegetable intake but a strong positive association with multiple adenomas. The study by Kunzmann and colleagues was a prospective cohort study that included 154, 952 participants aged 55-74 years and randomly selected from some studies in the United States between 1993 and 2001. The above findings suggested that fruit played a more protective role than vegetable intake in

the reduction of CRC incidence. The findings are significant to my dissertation project, as the destruction of farmlands, especially in the Niger Delta region of Nigeria, may decrease the availability of fruits and vegetables due to oil pollution and land degradation. The challenge is also possible in other regions, which are facing either constant draught or the Boko-Haram insurgence, as in the northern regions of Nigeria.

Age, Gene-Environmental Interaction, and Colorectal Cancer

Age as a non-modifiable risk factor has acted independently and interacted with other non-modifiable risk factors to inform CRC incidence and disparity. Gabriel et al. (2018) confirmed in their study that some biological factors differences influence the overall survival rates of patients with CRC between those aged 50 years and below and those aged above 50 years. In this study, Gabriel and colleagues queried the US National Cancer data set (NCDS) collected between 2004-2013 for patients with CRC. Included in this study were 870,030 CRC patients (488,121 colon, and 181, 909 rectal). The cases were stratified by age into two: Those 50 years and below and those above 50 years. Multivariate statistical analysis was used to identify risk factors associated with the overall survival of CRC patients. The result showed that cases aged 50 years and below had a higher proportion of pathological stages III and IV of CRC compared to those aged above 50 years. Disparities associated with the overall survival of both colon and rectal cancer showed that those aged 50 years and below had higher risk factors associated with the increase in mortality than those aged above 50 years. The authors concluded that there were differences in biological factors and the overall survival rates of patients with CRC due to age. Though this data is from a large nationwide database, the sampled population that

reflected only the white ethnic group/race will affect its external validity in other populations. Moore et al. (2018) research findings reinforced the findings of Gabriel et al. (2018). Moore et al. (2018) used a secondary data set from the California Cancer Data System between 1981-2013 to identify CRC incidence between those aged less than 50 years and those above 50 years of age. Multivariate logistic regression Statistical analysis included 182, 095 cases that satisfied the case definition and age risk factor. Results showed that about 50% of all valid cases were diagnosed at the advanced stage presentation, while 69.6% among the cases were those younger than 50 years. Based on the evidence available, the researchers concluded that younger patients aged less than 50 years were significantly more likely to be at risk of advanced CRC than those above 50 years of age. The above findings of the two studies, including two earlier consistent study findings, using the US population by Fairleyet et al. (2006) and Siegel et al. (2017), are significant in targeting CRC screening. This is because those above 50 years of age were thought to be at higher risk than the younger population, which informed the choice of the targeted population for CRC screening. The findings suggest that the use of age-adjusted cancer incidence, as in the case of my dissertation, will not allow age to confound the findings. The preceding is because CRC data from the cancer registries in Nigeria are provided as an age-adjusted CRC incidence data set. In the main, age is a non-modifiable risk factor in CRC disparities. Gene-environmental interaction is another possible non-modifiable risk factor, which is discussed below.

Epidemiological study findings suggested that genetic factors may be causal to both familial and sporadic CRC cases. Valle (2014) argued that familial CRC is associated with

genetic changes in high, moderate, and low-penetrance *susceptibility genes*. Evidence provided through the Genome-Wide Association Studies (GWAS) has also shown that some low-penetrance alleles are associated with an increase in CRC (Valle, 2014). Therefore, the accumulation of low-risk variants may explain, in part, the risks associated with the rise in CRC. GWA studies were only able to identify 40% CRC susceptibility loci. However, only a small percentage of heritability was explained by GWAS. The gene-environmental interaction between GWAS-recognized single-nucleotide polymorphism (SNP) and the environmental risk or protective factors for CRC have been used to account for the majority of the heritability factors (Song et al., 2019).

Valle (2014) performed a comparative review of genetically characterized and uncharacterized hereditary CRC syndrome for the low and moderate loci and variants using genome-wide associations and candidate-gene studies. The researchers posited that lifestyle and environmental factors interact with several multiple genetic risk factors to inform an increase in CRC incidence. Some other researches have shown that families with the fCRC-X gene have no germline mutation in the MMR genes. Therefore, the gene-silencing mechanism might also explain fCRC-X cases that the current mutation identification techniques cannot detect. Besides, candidate-gene studies have shown that the high-penetrance genes may be responsible for the incidence of uncharacterized familial CRC.

Shiao et al. (2018) conducted a study for assessing the gene-environmental interaction and prevention of CRC risks. The above study by Shiao and colleagues was a proof of concept study that included 108 participants (54 colons and 54 matched

family/friend controls) from a California Cancer Registry. Key genes in the *one-carbon metabolism pathway* and some environmental and lifestyle risk factors were all included in the study model. Shiao and colleagues used the three-explanatory family-based analysis to adjust for the sharing of genetic heritage. The final step in this study methodology was the inclusion of a generalized regression model to predict CRC risks. Based on the evidence provided by the statistical analysis, the authors concluded that the environmental and modifiable risk factors played a predictive role in the prevention of CRC using the gene-environmental interaction. The study is of significance to the dissertation understudy, as it included family members as an ecological unit in the methodology, which will reinforce the inclusion of regions as an ecological unit in the dissertation methodology. In a more current study, Song et al. (2019) evaluated gene-environmental interaction for CRC susceptibility loci using only cases and case-control studies. The above study design included 703 CRC cases and 1456 controls identified from the national cancer center in Korea between 2010-2013. The authors tested interactions between 31 GWAS-identified SNPs and 13 protective or risk factors of CRC. The logistic regression models were used to estimate gene-environmental interactions as it relates to CRC risks. This study's findings showed that SNP rs4444235 at 14q 22.2 interacted with the relationship between regular exercise and CRC risks. However, the study suggested that while the risk alleles of rs4444235 increased CRC risk in the regular exercising persons (OR=1.47, 95%CI: 1.02-2.10), it decreased the CRC risk in those non-exercising persons (OR=0.76, 95% CI: 0.62-0.94). The study also suggested a statistically significant association in the interaction between regular aspirin intake and the SNP rs2423279 at 20p12.3 (Song et al., 2019). Even

though the above study did not include all the CRC susceptibility loci identified by previous studies, its strengths lied in the large sample size and the use of controls. Gene-environmental interaction can be a risk factor in the incidence of CRC and a predictive factor in the prevention of CRC. Gene-environmental factors can also support the hypothesis of the role of life course epidemiology in CRC incidence, which is discussed below.

Life-Course Events and the Risk of Colorectal Cancer

The life-course epidemiology approach is the investigation of the long-term effects of exposures. According to Kuh and Shlomo (2004), the life-course approach includes biological, environmental, and psychosocial factors operating across the individual life or a cohort's life expectancies. The life-course approach to investigating chronic diseases risk acknowledges that every stage in life is relevant to health. It addresses the relationship between adverse health conditions over time and the risk of chronic diseases onset (Ben-Shlomo & Kuh, 2002). Some of the life-course epidemiological models include the critical period model (Ben-Shlomo & Kuh, 2002), the accumulation of risks model (Kuh & Shlomo, 2004), and the pathway or chain of risks model (Burton-Jeangros, 2015). Most life-course models used either the birth cohort or the prospective longitudinal studies (Blane, Netuveli, & Stone, 2007).

Clarke and Joshua (2017) examined early life exposure and adult cancer risks. The authors used a narrative review of evidence to describe the epidemiological evidence relating to early exposure to cancer risk factors and the adult onset of cancers. The life-course framework guided Clarke and colleague's literature search and review. Included in

their search were articles related to early-life exposures to environmental factors, including exposures during uterus, childhood, and adulthood stages in life. The review was conducted mainly using the MEDLINE searches for peer-reviewed articles written in the English language through July 2016. Clarke and Joshua concluded that there was evidence to suggest a positive relationship between early life exposures and cancer risks, which may inform primary prevention to reduce risk factors at the early stage of life development. The above study is significant to the dissertation project as it provided evidence-based data that sharpens biologically-based hypotheses, especially regarding time-period and cancer etiology.

Kim et al. (2018) used meta-analysis to establish a relationship between air pollution and all cancer mortality. Elsevier's PubMed, EMBASE, pharmacological, and biological databases were selected for the review, which included 30 cohorts (prospective and retrospective) studies. A random-effect statistical model was used to determine meta-estimates of each air pollutant. Based on the review of the evidence, Kim et. al. (2018) concluded that there was a strong association between long-term air pollution and cancer mortality. In particular, PM 2.5 was considered a risk factor for CRC, which reinforced the evidence provided by previous studies, especially the large epidemiological study by Turner et al. (2017).

Yamashita et al. (2018) discussed the embryonic origin of chronic diseases like CRC. The purpose of their study was to determine the prognostic value of embryonic origin of colon cancer liver metastases in patients undergoing resection after chemotherapy. The study included 725 patients with primary colon cancer who underwent hepatic resection

after pre-operative chemotherapy between 1990-2015. In addition, 252 patients with primary colon cancer who underwent resection were used as a control to validate the study. Yamashita et al. (2018) concluded that colon cancer was associated with pathologic response to chemotherapy and worse survival after resection. The emerging evidence from the literature suggested that the life-course approach to health is important in preventing chronic disease risks, like CRC, starting from pre-conception through adolescence and adulthood (Jacob et al., 2017). Besides, life-course would be important in determining causes of health disparity across regions (Jones et al., 2019). Therefore, an effective intervention to reduce the burden of chronic non-communicable diseases like CRC will require action at every life stage (Mikkelsen et al., 2019).

The research findings based on the life-course models justified the use of ecological study that uses group or ecological unit data rather than the individual risk factor to inform CRC risks in the population under study. The above argument is because the emerging evidence supported the notion that chronic disease is caused by the accumulation of risks or negative exposures across life stages. Also, using one or few exposures to assess the risk of CRC may be confounded by other risk factors not adjusted. Accounting for all the potential risk factors may also not be possible because some are mixtures of volatile organic compounds that may not have correlates for the biomarker detection (Peter et al., 2012).

Spatial Autocorrelation and Colorectal Cancer

Spatial autocorrelation in any disease outcome occurs when the value of a variable at a specified geographic location depends on its values at adjacent locations (Rosenberg,

2000). The occurrence of autocorrelation because of the contiguity of regions may be a source of violation of some statistical tests' basic assumption of the independence of occurrence of events (Daniel & Cross, 2013). Spatial autocorrelation may also be of significance in informing public health interventions to focus on populations most at risk of any disease of interest.

Al-Ahmadi and Al-Zahrani (2013) conducted a spatial autocorrelation analysis of various cancer incidence in Saudi Arabia. The authors used a secondary data set on various types of cancer collected between 1998 and 2004. Included in this study were 45,532 cancer cases (22,930 males and 22,602 females). Spatial autocorrelation analysis was performed for ten different distances starting from 5km to 500km. The global and local Moran's autocorrelation index were calculated to estimate any autocorrelation of the variable of interest and the possibility of clustering. The authors also carried out the '' Ordinary Least Squares'' (OLS) and the GWR regression analysis. The results of the study showed that different types of cancer exhibited different degrees of spatial autocorrelation at different cities and distances, including gender disparity, after Bonferroni corrections. For instance, male lung cancer and female breast cancer exhibited positive statistically significant spatial autocorrelation, indicating a clustered pattern. Although both the OLS and GWR regression analyses showed a positive statistically significant spatial association amongst different types of cancer in the population, the overall study results showed no statistically significant evidence to support any spatial autocorrelation of most cancers in the Saudi Arabia population. However, despite the weaknesses in the ecological study design used in this study, the presence of clusters through spatial analysis and associations

from the regression analyses can be useful in generating an etiologic hypothesis of cancer causation and spatial epidemiology of cancer in Saudi Arabia. Other studies described below provided further evidence to support spatial autocorrelation and CRC, which is of significance in one of the research hypotheses in this dissertation.

Li et al. (2017) specifically conducted a Spatio-temporal analysis of CRC incidence in Guangzhou, China. This spatial autocorrelation and retrospective spatial-temporal analyses included 14, 618 (8170 male and 6448 female) CRC cases with a crude incidence of 35.56/100,000 and an age-standardized incidence rate of 23.58/100,000 standard population. The data was collected between 2010 and 2014 from the Guangzhou cancer registry system. ‘’OpenGeoda’’ 1.2.0 software was used to determine both the global and local spatial autocorrelation indexes. The result showed a *global Moran index* of 0.527 (Z-11.06, $p<0.001$), suggesting spatial autocorrelation and cluster distribution pattern of CRC in the population. Thus, the results showed a spatial autocorrelation of CRC in Guangzhou, which is important in identifying populations with a high risk of CRC and the allocation of resources in CRC prevention. The use of spatial and Spatio-temporal analyses is a major strength, while the absence of covariate analysis is a major limitation in this study. The result of this study supports the research hypothesis of the second research question in this dissertation that there is a possibility of spatial autocorrelation of CRC amongst the six independent regions in Nigeria.

An earlier study focused on the retrospective spatial autocorrelation analysis of CRC in Kuala Lumpur was conducted by Shah et al. (2014). This study included a small sample size of 146 CRC cases (male 52.10%) collected from the Wilayah Persekutuan

heart center between 1995 and 2011. ArGis version 10.0 software and the SPSS version 19 were used to perform the statistical analysis of the data provided. The result showed no spatial autocorrelation of CRC in this population under investigation (*Moran's index*=0.007). However, the study revealed the clustering of CRC disease outcome (p<0.001, Z-score=-5.59). Even though the *Moran index* showed no spatial autocorrelation of CRC cases between adjacent neighbors, the presence of a statistically significant clustering pattern of CRC in Kuala Lumpur could be useful in targeting high-risk populations. The study findings also reinforced Al-Ahmadi and colleagues' earlier retrospective review findings that confirmed a statistically significant spatial clustering of some cancer incidence in Saudi Arabia without a commensurate spatial autocorrelation.

Mansori et al. (2018) used an ecological study design to conduct a spatial analysis of CRC incidence and the associated socioeconomic factors in the neighborhood of Tehran, Iran. The study included 2815 CRC cases aged 50 years and above. Data on CRC cases between 2008-2011 were obtained from the Tehran population-based cancer registry, while the data on socioeconomic variables were obtained from Equity Assessment Study conducted in Tehran. Both the spatial autocorrelation analysis and the Bayesian analysis using the Baseg-York-Mollie (BYM) model were conducted. The result showed that the *Moran autocorrelation Index* was statistically significant for all variables *(*p<0.05), while the Bayesian analysis indicated that socioeconomic variables were responsible for CRC disparity across the Tehran population and its neighbors. Thus, Marson and colleagues concluded that there was an inequality in the spatial distribution of CRC in Tehran and its neighborhoods, which is influenced by various socioeconomic factors. In an earlier similar

ecological-epidemiological study conducted in Massachusetts, the United States, Dechello and Sheeham (2007) confirmed a statistically significant autocorrelation of CRC in Massachusetts. The authors used data on CRC incidence (6360 males; 6628 females) and the proportion of late-stage in residents and collected between 1995 to 1998. The findings of spatial autocorrelation using ecological studies support this dissertation's study design and the second research question hypothesis of spatial autocorrelation of CRC within the six-independent regions in Nigeria under investigation.

These pieces of research evidence provided by Dechello and Sheeham (2007), Al-Ahmadi and Al-Zahrani (2013), Shah et al. (2014), Li et al. (2017), and Mansori et al. (2018) supported the hypothesis that spatial autocorrelation may be responsible for CRC disparity, including the spatial distribution of disease outcome that may be influenced by socioeconomic factors and location distances between neighbors.

In conclusion, there is evidence to support the possibility of spatial autocorrelation of CRC within Nigeria's regions that can be influenced by the socioeconomic variables of each region and location contiguity between regions. Information on the spatial distribution of diseases, such as CRC, is important in providing knowledge about the spatial dependence between contiguous locations and to plan for an effective differential and targeted CRC screening interventions and the allocation of health resources in favor of the most vulnerable population to the disease of interest.

Summary and Conclusions

The emerging evidence from the reviewed literature indicated that colorectal cancer is a chronic disease with many risk factors. These risk factors that account for CRC

disparity across populations included but were not limited to general environmental pollution, oil and gas pollution, gender, genetics, gene-environmental interaction, and life-course events, including spatial autocorrelation between contiguous locations. Unfortunately, most of the reviewed literature failed to anchor their research premise on any theoretical foundation explicitly. However, the implied theoretical framework included psychosocial/Ecosocial (Canchola et al., 2017; Doubbeni et al., 2012), socioeconomic determinants (Lpez-Abente et al., 2012; Kikuchi et al., 2017; Zhang et al., 2018), social determinants (Hurtado et al., 2015), and life-course accumulation of risks due to various exposures (Kuh & Shlomo, 2004; Ben-Shlomo & Kuh, 2002; Clark & Joshua, 2017; Burton-Jeangros, 2015), as shown in Table 1.

Table 1

The Implied Theoretical Foundation of Some of the Reviewed Articles

Theoretical foundation	Authors
Psychosocial/Ecosocial	Kim et al. (2010), Nnagu et al. (2016), Kruk et al. (2019).
Socioeconomic disparity/Health determinants	Lopez-Abente et al. (2012), Yuan et al. (2018)
Life-course/Cumulative risk or webs of causation	Kuh and Shlomo (2004), Ben-Shlomo and Kuh (2002), Clark and Joshua (2017), Burton-Jeangros (2015)

Study methods or designs used in the reviewed literature included prospective cohort studies (36%), retrospective cohort studies (9%), case-control studies (11.5%)),

cross-sectional studies (5.7%), meta-analysis, and review of the evidence (21%), ecological studies (5.7%), and others (15.4%), as shown in Table 2. Life-course models were also included in a few reviewed literature, especially those based on the accumulation of risk and life trajectory effects. The reviewed ecological studies of interest to my dissertation under consideration included those conducted by Lopez-Abente et al. (2012), Orleswka et al. (2017), and Ayuso et al. (2020). It is important to note that about 75% of all the observational studies in this literature review used the secondary data set, while the remaining 25% used primary data to answer their research questions. Secondary data sets, especially from the US population, were prominent in the reviewed literature. The data from the National Cancer System Registries (NCSR) in Nigeria was not used by any of the studies to answer research questions related to living in oil-polluted regions and the risk of excess CRC incidence. What is not known in this discipline or study relating to the topic of this dissertation is the use of group data and the influence of spatial autocorrelation analysis to inform the relationship between CRC putative risk factors, including oil pollution, to inform CRC disparity in the population studied.

Table 2

Study Designs Used in Some of the Reviewed Literature

Study Methods/Design	No of Reviewed Articles or Literature
Prospective Cohort	19
Retrospective Cohort	5
Case-control	6
Cross-sectional	3

Systemic literature review	10
Ecological studies	3
Others, including spatial autocorrelation	12

Gap in Literature

None of the reviewed literature used any study design, including the ecological study model, to establish the relationship between CRC risk and living in oil pollution in the Nigerian population. The above constituted the gap in knowledge and literature in this domain. Therefore, this study aimed to determine an association between living in an oil-producing region and the risks of excess CRC incidence in Nigeria using an ecological study. Filing this gap in literature and knowledge is important because the outcome of this research will inform an effective public health intervention and practice in environmental pollution management to reduce any existing colorectal cancer disparity across the six regions or geopolitical zones in Nigeria. The ecological study is appropriate as the study is interested more in the contextual effect of environmental risk factors on the population under investigation. Addressing the perceived disparity in health outcomes, such as colorectal cancer because of the differential exposure to environmental pollution, will minimize agitations and promote social change in the population under study (Moses & Olaniyi, 2017). The above is because ecological study design in this study is premised on the assumption that differences in exposure between groups are usually bigger than measurement at the individual level, and ecological studies are better to inform policy changes at the population level (Stevenson & McClure, 2005). The study may also

contribute to the ongoing scholarly inquiry on the best way to allocate resources to the various regions based on their exposure to health challenging environmental risk factors. Besides, the use of the six independent regions in Nigeria as ecological units justified the use of cumulative exposures against the individual risk factors or exposures. The research methodology in this study addressed the gap in literature. The research design and methodology, including the population description, sampling methods, exposure and outcome variable measures, secondary data set analysis, data analysis plan, possible threats to validity, and the concluding summary are reported in the research methods, which is the Chapter 3 of this dissertation document.

Chapter 3: Research Method

Introduction

This study's primary purpose was to investigate whether living in Nigeria's oil-producing regions is responsible for any excess CRC incidence. I also investigated the spatial autocorrelation of CRC outcomes between contiguous regions in Nigeria. In addition, I examined the possible relationship between gender and CRC disparity as a potential measure of occupational hazards and risks within the population under study (Scarselli et al., 2018.

In this study, I analyzed secondary data sets from the NSCR, the IQAir real-time exposure data, and the 2018 Nigerian Demographic and Health Survey (NDHS) to answer the RQs using a quantitative, correlational, ecological research approach. The ecological study approach was chosen because differences in environmental exposures are usually bigger within groups than at the individual levels (Stevenson & McClure, 2005). Ecological studies are used to monitor population health for an efficient public health intervention (Levin, 2003). The ecological study design was more appropriate than one involving individual-level data because it is interested in the contextual effect of environmental risk factors on the population under investigation. Evidence from ecological studies is useful in implementing public health policies and interventions (Stevenson & McClure, 2005), such as those expected from this study's outcome.

In this chapter, I present the research design and the rationale and research methodology for the study. The Methodology section includes information on the population, sampling frame, sampling method, sampling procedure, data collection,

secondary data, statistical analysis plan, threats to validity, and ethical issues. This chapter ends with a summary and a transition to the next chapter.

Research Design and Rationale

This study's dependent variable was the age-standardized CRC incidence across the six independent regions, a quantitative continuous variable measured at the scale level. The independent variable for the first RQ was region, as represented by the regional exposure values for PM 2.5, which was a proxy pollution candidate for the risk factors for CRC in the populations under study. I operationalized contiguity between regions and gender as independent variables for the second and the third RQs. Covariates that were included in this study's statistical model were the Gini Income Index (World Bank, 2003), wealth quintiles or index, education, and access to health insurance. Although the Gini Index measured the level of income inequality within each region, the wealth index measured the household consumer goods at the disposal of each family unit covered in the survey. Each household was allocated scores, which were derived using the principal component analysis, which is a method used to reduce large data sets by transforming a large set of variables into smaller ones without losing the main information of the original large data set (Jolliffer & Gadima, 2016). The education variable was measured as the percentage of those who attained the highest education level, while access to health insurance was measured as the percentage of the population in each region that had access to at least one type of health insurance coverage. I measured all the covariates at the scale levels.

The study involved the use of group data to answer the RQs. The quantitative ecological study design was informed by the RQs and the theoretical construct, a derivative of the theoretical framework. I conducted an ecological study because the main RQ centered on CRC disparities among the six independent regions in Nigeria and not at the individual level. According to Wakefield (2007), researchers conducting ecological studies use data aggregated over groups instead of individual data. Data on the individual measurement of the potential environmental exposures are not readily available for the population under investigation. Evidence suggests that the measurement of exposures at the individual level in a defined geographical region does not usually differ but does differ significantly from region to region (Hatch & Thomas, 1993).

Further, some of the exposures associated with CRC may not have correlates, especially the complex mixtures from oil and other environmental pollutants or exposures like noise and crude oil. Hence, measurement at the individual level of biomarkers may be challenging, if not impossible (Peters et al., 2012). I was interested in estimating the cumulative effect from ecological variables, including those for which there is no correlate or biomarker at the individual level. The use of an ecological study approach was also informed by the knowledge that chronic diseases like CRC can be a life course event. Godfrey and Baker (2000) hypothesized that chronic diseases like CRC might occur because of the accumulation of so many socioenvironmental factors, including ecosocial and psychosocial factors, such as stress due to inequalities, gene-environmental interaction, and other socioenvironmental factors related to the fetal or developmental origins of adult disease (Calkins & Devaskar, 2011). I also designed the study to provide data at the group

or regional level to support the ongoing scholarly inquiry (e.g., Nnaemeka (2020) on the impact of oil pollution on health outcomes among communities living in Nigeria's oil-producing region. All the data on covariates are available at the regional level (ecological unit) from the NDHS data set. Hence, an ecological study design was appropriate.

The concept of the exposome is assumed in the theoretical framework. According to Stingone et al. (2017), the exposome is the lifelong integration of all exposures that may influence human health**.** Based on the concept of the exposome and theoretical framework, I used the six independent regions or groups as the independent variable, while the dependent variable was the aggregate data or the mean of the CRC incidence from all the individuals in all the states that make up each region. Aggregate ecological variables are a composite measure or summary derived from values collected from the individuals. The assumption is that only the individual gets the disease, but the cancer or cancer incidence rate is a summary for all the individuals in the population or region (Stingone et al., 2017). Even though ecological studies do not necessarily confer causality per se, the use of both the descriptive and inferential statistics in the secondary data analysis may provide a clue to causality, including the spatiotemporal evidence of the relationship between the variables. The major constraint in this quantitative, correlational ecological study design was access to the secondary data. This is because the NSCR secondary data is prepared by each cancer registry at each state or region and forwarded to the Federal Ministry of Health. The Federal Ministry of Health is responsible for data access contracts with prospective consumers for research. This process is time-consuming and may cost some money in motivating action by the relevant stakeholders across the value chain before the release of

data. Another major limitation of the ecological study design is the concept of ecological fallacy, which occurs when group data are used as proxies for individual-level data (Schwartz, 1994). I describe the ecological fallacy in detail in the Data Analysis Plan section of this chapter.

Methodology

This study was a secondary analysis using deidentified data from the NSCR, which is domiciled at and managed by the Federal Ministry of Health, and the pollutant exposure data from the ÍQAir global air monitoring company. Data on covariates came from the 2018 NDHS. The study population included mainly persons living in each region covered by the NSCR data set who met the predefined case definition. I developed the RQs before searching for the appropriate data set to answer them (General Medicine, 2012), in order to avoid bias from data dredging that usually accompanies a prior knowledge of data set variables before stating the research question (Davis-Kean et al., 2015). This approach also allowed me to apply theoretical knowledge and conceptual skills based on the a priori hypotheses to utilize existing data to address the RQs (Johnson, 2014). The NSCR data were published by the Federal Ministry of Health. The data were collected from 2009 through 2013. The data set included an age-adjusted cancer incidence rate and consisted of 18 registries that submitted data to NSCR. I extracted only the data for CRC from this data set. The NSCR was established in 2009 by the federal ministry of health to coordinate and provide the various cancer incidence data across all regions in Nigeria (Jedy-Agba et al., 2016). According to Jedy-Agba et al. (2016), NSCR is a data set that contains age-adjusted

cancer incidence in Nigeria to provide the policy makers with evidence-based data on cancer epidemiology in Nigeria.

To ensure database integrity and validity, policy makers have mandated regular audits and reporting of the NSCR. This data set's strengths are its noncomplex nature with uniform case definition across all regional cancer registries, including the coverage of more than 85% of the population of interest (Jedy-Agba et al., 2016). Based on my review of the literature, no other researcher has used this data set to answer the same RQ under consideration in this population. Therefore, this data set's use in this inquiry increased this study's novelty and significance (see Smith et al., 2013). The large sample size of NSCR provided greater precision of parameter estimates and allows for generalization of the outcomes of this study to other populations.

The 'IQAir' data creators use air visual sensors to provide real-time PM 2.5 air pollution data through the organization's air quality monitoring information platform (IQAir, 2021). The organization uses an air quality app and website to provide a centralized platform for global and hyper-local air quality information in real-time. The app and website operate through the organization's ground-based stations and use artificial intelligence to calibrate and validate air quality monitoring stations. Data are collected at monitoring stations and grouped into cities or settlements, which are further aggregated into states, regions, and the entire country.

The Nigerian National Population Commission gathered the 2018 NDHS in collaboration with the National Malaria Elimination Program of the Federal Ministry of Health. The NDHS project is sponsored by the World Health Organization, United States

Agency for International Development, Bill and Melinda Gates Foundation, and the Global Fund (National Population Commission, 2019). It is a survey that is aimed at providing up-to-date data on key demographic and health indicators in Nigeria (National Population Commission, 2019).

I obtained the data dictionary or codebook of the NSCR data set through a literature review after defining the RQ. The data dictionary revealed that the dependent variable is a continuous variable with a scale level of measurement. Also, I assessed data on pollution exposure online as part of the literature review. The data on PM 2.5 are reported as counts with a scale level of measurement. The covariates data from the NDHS were mainly continuous variables measured at the scale levels. Kornegay and Segal (2013) argued that one of the key considerations for selecting a data set is its ability to disclose the appropriate variables and their measurement levels.

Population

The study targeted colorectal cancer incidence across the six regions in Nigeria, as shown in figure 1, including gender variations and the influence of a possible spatial autocorrelation between neighboring or adjacent regions or zones. Each region or zone is made up of distinct ethnic groups with different cultures and geography. Each region is involved in a different form of agriculture and commerce. While the entire northern regions are known for a relatively large scale of agriculture, the southwest controls commercial activities like banking. The southeast is leading in trade and commerce, while the southsouth is known mainly for its oil deposits, which account for about 85% of Nigeria's total foreign earnings. According to the article published in the African Population Studies

by Ikoba et al. (2018), each region or zone represents the following percentage of the total population, as of the 2006 last official census: northwest (25.58%), southwest (19.74%), southsouth (14.99%), northcentral (14.51%), northeast (13.52%), and the southeast (11.68%), as seen in Table 3. Environmental pollution due to diesel, gasoline, bush burning, and automobile noises are common to almost all Nigeria regions. However, environmental pollution from oil and petrochemicals is prevalent mainly in the southsouth region because of oil exploration and exploitation by multinational oil companies like Shell, Chevron, and Mobil.

States in the southsouth region, which are part of the entire Niger Delta region, are contiguous to some states in other regions like the southeast and southwest, which may affect the spatial distribution of health outcomes, such as CRC, because of the possibility of spatial autocorrelation between adjacent regions. According to Rosenberg (2000), Spatial autocorrelation is a situation in which the value of a variable at a specified geographic location depends on its values at adjacent locations. This neighborhood effect can lead to confounding due to spatial clustering and autocorrelation between the adjacent regions, as seen in Figure 1. The above-described possibility of spatial autocorrelation is consistent with the medieval or ancient ‘’first law” of geography, which posits that “everything is related to everything else, but near things are more related than distant things’’ (Al-Ahmadi & Al-Zahrani, 2013). Therefore, based on the research hypotheses of this study, it is expected that the people living in the southsouth region should have the highest CRC incidence, followed by the people living in the southeast and southwest

regions because of their location contiguity to the southsouth region when compared to all the regions in the northern part of Nigeria.

Table 3

Different Regions in Nigeria and Their Ethnicity and Percentage Contribution to Total Population

Regions	States	Ethnicity/Socioc ultural group	% of total Populati on.	% of Female Populat ion	% of Male Populati on.
Northcen tral	Benue, Niger, Kogi, Nassarawa, Kwara, and Plateau	Mixed ethnic groups.	14.51	14.7	14.9
Northeas t	Adamawa, Bauchi, Gombe, Borno, Taraba, and Yobe	Kanuri and Hausas.	13.52	17.9	17.9
Northwe st	Kano, Kaduna, Sokoto, Zamfara, Jigawa, Kebbi, and Katsina	Mainly Hausas and Fulani	25.58	27.5	25.7
Southeas t	Abia, Anambra, Ebonyi, Enugu, and Imo	Mainly the Ibos	11.68	9.7	8.8

Southsouth	Akwa Ibom, Bayelsa, Cross River, Delta, Edo, and Rivers	Mixed ethnic groups of Delta Region	14.99	17.6	19.0
Southwest	Ekiti, Lagos, Ogun, Ondo, Oyo, and Osun.	Mainly the Yoruba	19.74	12.6	13.7

Sampling and Sampling Procedures

In studies where it is difficult to access data for the entire population, it is always statistically prudent to select a representative sample. Such samples should be selected in a way that inferences can be made to the entire population based on the samples' data without any bias. The sample was selected through a stratified sampling technique from only the 18 registries across all the six regions that submitted data to the NSCR questionnaire in 2009.

A stratified simple random sampling was applied in the selection from the 18 registries that included mainly the population-based registries and a few hospital-based cancer registries. Stratified sampling was appropriate because dividing the entire population into sub-groups (regions) or strata ensured that the entire population was represented. It was also appropriate because the research question is trying to draw a conclusion from the sub-groups. In this probability sampling, the total number of cancer registries across the country was first stratified into states and regions. A simple random sampling was then conducted using a random number so that the selected registries were representative of each region's population of cases and ensured that there is no double-

counting or inflation of cases in favor of any region. The sampling frame therefore, included representatives of all the six regions at the national level. All the hospital-based cancer registries were excluded except in cases where there were no population-based cancer registries in some regions when making decision on the sampling frame. The total sample size was targeted to be enough to power the study to about 85%.

In this study, a power analysis using the ‘’G Power’’ software (Faul et al., 2007) was performed to determine the appropriate sample size to power the research to ensure internal and external validity and avoid type 1 error. As the statistical test for the main research question is the one-way ANOVA, large sample size was needed to achieve an approximately normal distribution of the dependent or outcome variable so that the statistical test is conducted without any violation of the proposed ANOVA test. The ‘’G Power’’ statistical software was used to calculate the sample size, given an expected 85% power (the probability of not committing type 2 error), an alpha value of 0.05, and an effect size of 0.30, which was selected based on literature reviews (This effect size is the minimum deviation from the null hypothesis that I hope to detect). Selecting the fixed effect, omnibus, and one-way ANOVA test in the ‘’G’’ Power software, and inputting the number of groups as six (six independent regions), the approximate total number of samples needed to achieve a power of 85% in this study was found to be 168.

Procedures for Recruitment, Participation, and Data Collection

The secondary data from the ‘’National System of Cancer Registries’’ (NSCR), IQAir, and the 2018 NDHS data were used to answer the research questions. The NSCR secondary data set for this analysis was collected between 2009 and 2013, while the NDHS

was collected in 2018. The NSCR presently consists of 14 population-based and 20 hospital based-registries. However, only data from the 18 registries that submitted data to NSCR were analyzed. The NSCR was established in 2009, after several failed attempts to regulate cancer data reporting in Nigeria since 1960. The NSCR was established through a collaborative effort between the Federal Ministry of Health of Nigeria, the Society of Oncology and Cancer Research in Nigeria, and the Institute of Human Virology in Nigeria. According to Jedy-Agba et al. (2016), NSCR was established to provide cancer incidence data from all Nigeria regions. The NSCR project was supported by the Fogarty International Center and the National Cancer Institute, and the National Institute of Health (NIH) of the United States, including the Marlene and Stewart Greenebaum Cancer Center and the University of Maryland School of Medicine. The main objective of the NSCR was to coordinate the activities of the cancer registries across the country and to generate aggregate national cancer incidence, treatment, and survival data, including the dissemination of evidence-based cancer data to relevant government agencies for informed policy interventions and appropriate resource allocation in Nigeria. In addition, the NSCR was established to provide training, capacity development, mentoring, technical and scientific support to cancer registries in Nigeria to enable them to attain population-based cancer registry status and generate high-quality cancer incidence, treatment, and survival data for the country (Jedy-Agba et al., 2016). The NSCR data set agreement link is available to researchers and the public on the NSCR website. However, access to the main data file was only through a data agreement with the management of the data set domiciled within the Federal Ministry of Health. The agreement form was available online.

Submission of an approved research proposal by the prospective consumer of the data set was required to give effect to the data-sharing agreement. However, the request for the submission of an approved research proposal before the release of the data set was perhaps to ensure that data are collected and consumed for the purpose intended.

The NSCR data within the period under review (2009-2013) were collected from 18 registries that responded to the data submission questionnaire requested by the NSCR. Sixteen registries did not respond to the questionnaire. Therefore, no data from the non-responders was included in the 2009-2013 NSCR data set. The collated data were checked for consistency and completeness, and other data quality parameters, using the CanReg4 & 5 Software, depending on the version that was available in each of the registries that submitted data (Jedy-Agba et al., 2016)). Case definition and diagnosis were mutually agreed upon, while the age-standardized incidence rates (ASRs) were calculated for each of the population-based cancer registries. Each population-based cancer registry's quality was assessed using the indices of completeness, validity, and timeliness.

The data collection methodology used in this data set is consistent with the global best practices for cancer registries. Descriptive statistics based on gender were applied, and the basis of diagnosis for the outcome variable was clearly defined. The data set validity and integrity were further challenged by evaluating the measurement of variables of interest, possible measurement errors, sample size and sampling frame, and the missing data. In my opinion, the NSCR data set, as presently constituted, satisfies the minimum requirements for a reliable and valid data set. This is because, according to Boo and Froelicher (2013), the major threats to the validity and reliability of any secondary data set

are premised on poor methods of data collection, missing data, measurement error, and the sample frame selection. However, my poor understanding of the data entry process and the lack of my direct involvement in this data collection process posed a potential source of limitation to the use of this data set to answer my research question (Clarke and Cossette. 2000). All information collected from this data set will be kept confidential.

The independent variable data was obtained online in real-time from the ‘’IQAir’’ global air quality monitoring ground-based stations. IQAir uses data on PM 2.5 as a representative measure of environmental pollution. The preceding is because PM 2.5 is widely regarded as the pollution with the most health-related effect, such as CRC, compared with other commonly measured air pollutants (Viatte et al., 202). IQAir uses air quality sensors with an internally-developed advanced light-scattering laser sensor, which is about the most accurate and sensitive means to measure PM 2.5 exposure concentrations. Each station monitor is carefully tested and calibrated in a factory setting. The device is exposed to varying pollution levels using a controlled pollution environment calibrated via a computer system, which ensures high precision to reference monitors and low intra-model variability. IQAir used the US air quality index (US AQI) for its value measurement. The US AQI is the most commonly recognized AQI system available and is communicated daily to the public (Kevin et al., 2020). The US AQI computes pollutant concentration into a color-coded scale of 0-500 to represent the level of associated health risks. It is stratified into-good, moderate, unhealthy for sensitive groups, unhealthy, very unhealthy, and hazardous. The ‘’Good range’’ value is < 12ug/m3, which is slightly higher than the World Health Organization air quality guideline value of < 10ug/m3.

For the data on covariates, the NDHS data was used. The NDHS data were collected between 14th August and 29th December 2018. It used a two-stage stratified sampling technique that identified 74 sampling strata and 1,400 enumeration areas. Data were collected from male and female respondents aged 15 to 59 years using the computer assisted personal interviews with three questionnaires that also included questionnaires on biomarker data collection. Data entry and editing were performed using the CSPro software package. Out of the 40,666 occupied households selected, 40,427 households were successfully interviewed, representing a 99% response rate. The data set was accessed through the internet, as it is available online for public consumption.

Instrumentation and Operationalization of Constructs

The data set for CRC incidence is updated every four years to monitor the various cancer incidences across the six regions or geopolitical zones in Nigeria. The NSCR monitors data input quality from the various participating registries through training and regular data and processes audits.

The variables that I used in this study were operationalized before the analysis, and I only had access to the NSCR data set after approval from the Walden University's Institutional Review Board (IRB). The study's dependent variable was the CRC incidence, which was operationally defined as the sum of all the age-adjusted CRC incidences from the states that make up each region. To reduce the potential confounding due to age-distributions of the population, the NSCR used a weighted average of the age-specific rates derived from the crude incidence rates between 2009-2013, where the weights represented the proportion of persons in the corresponding age group of the standard world population

(per 100,000 persons). The 2013 census figures in Nigeria were used as the base for calculating the age-specific rates. The Data Set Codebook from my literature review showed that the 'International Disease Classification'' (ICD-O) code for the CRC outcome variable is C18-20. The dependent variable (ASR) is a continuous variable with a scale level of measurement. The study's independent variable is the six independent groups or regions or geopolitical zones in Nigeria, represented by the real-time PM 2.5 pollution values across all Nigeria's major cities. It was operationalized as a proxy pollutant candidate for all the common pollution exposure putative risk factors responsible for CRC incidence in all the regions being investigated. The exposure variable was operationalized as a categorical variable with six-independent groups (levels) and measured at the nominal level to answer the main research question.

The dependent variable involved data manipulation to integrate all the incident cases in all the states that make up each region into one variable heading. Accordingly, all the incident cases from the states of each region were merged. Since the data files from the different regional registries had the same variables, data appending was the most appropriate data merging technique. Merging files is very important in data analysis as it allows users to merge two existing SPSS files (with extension. sav) by adding cases or adding variables. Appending two data sets requires that both should have variables with the same name (Lee, Famoye, Shelden, & Brown, n.d.a.). In the statistical analysis for this study, no new variable was created for the dependent variable. However, a new variable labeled as ''Regions'' was created with six-independent groups: southeast, southsouth, southwest, northeast, northcentral, and northwest, respectively. Other new variables

included the data on covariates from the NDHS data set. The data from the NDHS were available as group data, such as the main independent variable ‘’Region’’ (northcentral, northeast, northwest, southeast, southsouth, and southwest).

Outcome Variable Measure

In this study, the outcome variable is the CRC incidence measured as age-standardized incidence rates. The case definition across all registries is determined by clinical diagnosis, cytology, and history of metastases. It is reported as the age-adjusted incidence. It is a continuous variable with a scale level of measurement. All CRC cases not meeting the case definition were excluded.

Exposure Variable Measures

This study’s exposure variable is the level of PM 2.5 pollution used as a proxy for the risk factors for CRC incidence across the six regions in Nigeria (southeast, southwest, southsouth, northeast, northcentral, and northwest). PM 2.5 pollutant is chosen as a proxy pollutant candidate because evidence from the literature review suggested that PM 2.5 is a major risk factor for CRC (Ethan et al., 2020). PM 2.5 can penetrate different organs and the bloodstreams and evoke poor health outcomes, which may be largely because of its small particle size. Also, NO2, which is reputed as a major single risk factor for CRC in a large epidemiological study (Turner et al., 2017), is regarded as a component of particulate matter, such as PM 2.5 (Viatte et al., 2020). Fann et al. (2018) confirmed in their study that PM 2.5 is a common pollutant in the oil and gas industry, including pollution from motor vehicle oxidation products and bush burning that are common in both oil-producing and non-oil producing regions. Therefore, the choice of PM 2.5 as a proxy risk factor for CRC

for the six regions is well informed. Data for each city covering up to the local councils are aggregated for each state and then for each region and country. PM 2.5 is measured in ug/m3 or air quality index (AQI). It is a count variable with a scale level of measurement.

Apart from the pollution exposures represented by the PM 2.5 AQI values, each region is assumed to have its own cumulative exposures to other socio-environmental factors, such as stress and socioeconomic factors, including the nonmodifiable risk factors and the life-course effect over time. The exposure variable was operationalized as a categorical variable with a nominal level of measurement in answering the main research question.

Covariates

Data on covariates was obtained from the 2018 NDHS. Details on the covariates that were included in this study, including their definitions and measurement levels, are described in the research design and rationale section of this dissertation document.

Data Analysis Plan

Screening and Cleaning of Data Set

SPSS software version 27 (Antonius, 2013) was used to analyze the data emerging from this study. Cleaning of the NSCR, NDHS, and the environmental exposure data sets was performed before any data analysis to ensure that missing data, skip patterns, including any significant outliers were addressed. Cleaning also involved the examination of the non-response rate and the normality of distribution of the dependent variable in line with the ‘’Central Limit Theorem (CLT). According to Abdulwahab et al. (2011), cleaning and

screening data allow researchers to avoid violations of the proposed statistical tests' assumptions.

It is important to note that this study is vulnerable to ''Ecological Fallacy''. The ecological fallacy occurs when exposure measurements based on the group-data are used to make inferences at the individual level (Diez Roux, 2000). While aggregating data at the group level can reduce measurement errors, the use of group data as proxies to make inferences at the individual level may confound the result (Schwartz, 1994). However, no data analysis plan is made in this study to adjust for the ecological fallacy. The reason above is because this study's main purpose was to use group data at regional levels to confirm any CRC disparity and to inform differential public health interventions at the regional or geo-political zone level. Data at the group level is better to inform public health interventions than individual level data (Stevenson & McClure, 2005). The use of ecological regression models (Gilman & Park, 2001) and multi-level studies as means of accounting for the ecological fallacy would be recommended in this study for future research in this knowledge domain. The use of multi-level studies (group and individual-level data) requires measurements at the individual level, which is not available in this study's exposure data.

To ensure accurate data analysis for research question's variables, exposure and outcome variables and their levels of measurement were identified, recorded, and manipulated. Descriptive statistics were presented using tables and charts. At the same time, *p*-values and 95% confidence intervals (CIs) were calculated for the inferential statistics, which mainly included analysis of variance (ANOVA), *Global Moran*

autocorrelation index, *t*-tests, and the *multiple linear regression* analyses covering the three RQs and the covariates analysis.

Statistical Tests for the Research Variables

Both the descriptive and inferential statistics were calculated across the research questions and the variables of interest. While descriptive statistics enabled me to present the data in a more meaningful way, which allows easy interpretation of the data, the inferential statistics allowed me to make an informed decision on association measures, including any possibility of causal inference (Daniel & Cross, 2013). The types of statistical tests were largely informed by the variable types and their levels of measurement. The RQs and hypotheses, including the various variables and their level of measurement in this study, are described below:

RQ1: Is there a statistically significant relationship between living in the oil-polluted environment or region and colorectal cancer incidence in Nigeria? The research question examined a statistically significant relationship between living in an oil-polluted region and colorectal cancer incidence in Nigeria.

The independent variable is ‘’Region’’ represented by the Particulate Matter (PM 2.5) levels of pollution across the six independent regions, while the dependent variable is the ‘’age-standardized colorectal cancer rates’’. The dependent variable is a continuous variable with a scale level of measurement, while the independent variable is a categorical variable with a nominal level of measurement.

H01: There is no statistically significant relationship between living in the oil-polluted environment or region and colorectal cancer incidence in Nigeria

HA1*:* There is a statistically significant relationship between living in the oil-polluted environment or region and colorectal cancer incidence in Nigeria.

RQ2: Is there a statistically significant relationship between the contiguity of the six regions in Nigeria and the spatial autocorrelation of CRC incidence? The research question examined statistically significant spatial autocorrelation of CRC among the six regions in Nigeria.

The independent variable is ‘’Geography ‘’, represented by the ‘’contiguous regions,’’ while the dependent variable is ‘’colorectal cancer incidence rates’’. The independent variable is a categorical variable with six independent levels and a nominal level of measurement, while the dependent variable is a continuous variable with a scale level of measurement.

H02: There is no statistically significant relationship between the contiguity of the six regions in Nigeria and the spatial autocorrelation of CRC incidence.

HA2: There is a statistically significant relationship between the contiguity of the six regions in Nigeria and the spatial autocorrelation of CRC incidence.

RQ3: Is there a statistically significant relationship between gender and CRC incidence within and between Nigeria’s various regions after adjusting for covariates, such as education, wealth, Gini Index, and access to health insurance? The research question examined if there is a statistically significant relationship between gender and CRC incidence in Nigeria.

The independent variable is ‘’Gender’’, while the dependent variable is ‘’colorectal cancer incidence rates’’. The independent variable is a dichotomous categorical variable

with a nominal level of measurement, while the dependent variable is a continuous variable with a scale level of measurement.

HO3: There is no statistically significant relationship between gender and CRC incidence within and between Nigeria's various regions after adjusting for covariates, such as education, wealth, Gini Index, and access to health insurance.

HA3: There is a statistically significant relationship between gender and CRC incidence within and between Nigeria's various regions after adjusting for covariates, such as education, wealth, Gini Index, and access to health insurance

Descriptive statistics were calculated for both the independent and dependent variables to determine central tendency and dispersion measures. The descriptive statistics enabled me to determine if the statistical tests were conducted in violation of the proposed inferential statistical tests' assumptions. Some of the key descriptive statistics of interest in this study are the mean, median, standard deviation, and variance. Visual descriptive statistics included a bar chart and histogram. The histogram provided an indication as to whether or not the dependent variable is approximately normally distributed to avoid violations of the assumptions of statistical tests like ANOVA, Multiple linear regression, and *t*-tests, which are the main inferential statistical tests that was included in this study to answer most of the research questions. The choice of these statistical tests was largely informed by the types of variables and their levels of measurement.

In the first RQ, which is the main research question in this study, ''Region'' is identified as the independent variable and age-adjusted CRC incidence as the dependent variable. While the independent variable is a categorical variable with six independent

groups and a nominal level of measurement, the dependent variable is a continuous variable with a scale level of measurement. Based on these variable types and measurement levels, the appropriate inferential statistical test was the one-way ANOVA test. According to Daniel and Cross (2013), the One-way ANOVA is the appropriate statistical test when the dependent variable is continuous, and the independent variable is categorical, with more than two independent groups. The One-Way ANOVA compared the ‘’means’’ of the outcome variable across the six regions. A posthoc test was performed, and the initial omnibus ANOVA test informed the type of posthoc test (Tukey or Bonferroni). The ANOVA test revealed any statistically significant differences in the means of CRC incidence between those living in oil-producing areas (southsouth region) and the rest of the regions in Nigeria. The measure of association in this study was the F value in the ANOVA test. The ANOVA inferential statistics used two-tail tests with 95% interval limits and a p-value of significance of $p<0.05$. The analysis of covariate (*ANCOVA*) was also used to adjust for the influence of covariates like gender in the association. The 0.05 pre-determined alpha value is the level of risk that I am willing to take to reject the null hypothesis when I should not do so.

In the second RQ dealing with spatial autocorrelation, the ‘’Global Moran index autocorrelation inferential statistical analysis was employed (Li et al.,2017; Mansori et al., 2018; Shah et al., 2014). In this spatial analysis, it was assumed that CRC incidence in one region is not dependent on another contiguous region in space. The Global Moran I provided information on whether or not there was any form of dependence and preliminarily establish any evidence of clustering, dispersion, or randomness of CRC

incidence between contiguous regions. The ‘’Global Moran Index (*I)* spatial statistics was calculated to test for any spatial autocorrelation in this relationship between the outcome variable (CRC) across the various regions. Spatial correlation occurs if the outcome variable in some regions is influenced by the outcome variable in other regions, which are contiguous. In other words, spatial autocorrelation occurs when there is a spatial dependence of outcome variables between contiguous states or regions. The above is consistent with the Tobler's first law of geography that ‘’everything is related to everything else, but near things are more related than distant things’’ (Al-Ahmadi & Al-Zahrani, 2013). Spatial autocorrelation may be a source of confounding in this ecological study. According to Rosenberg (2000), the Moran Index *(I*) usually has a value that ranges between -1 to +1 (-1 value shows a negative autocorrelation, and +1 shows a positive autocorrelation or dependence, while a ‘’zero’’ value indicates that there is no autocorrelation, meaning that values of CRC incidence for contiguous (adjacent) regions do not depend on each other). A positive autocorrelation may be a source of confounding in this ecological study. This is because the CRC cases between adjacent regions depend on each other (autocorrelation).

Global Moran Index $$I = \frac{N}{W}\frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

where Nis the number of spatial units indexed by i and j; x is the variable of interest; $\bar{\mathbf{X}}$ is the mean of x; w_ij is a matrix of spatial weights with zeroes on the diagonal (i.e., $w_{ii} = 0$); and W is the sum of all w_ij.

In the third RQ that involved the determination of any possible gender disparity in CRC incidence outcome across and within regions, the *t*-test inferential statistical analysis was employed. The *t*-test was appropriate because the dependent variable is a continuous variable with a scale level of measurement, while the independent variable (gender) is a dichotomous categorical variable with a nominal level of measurement (Daniel & Cross, 2013). Multiple linear regression statistical test was also used to adjust for covariates like smoking, wealth, and education in determining their possible confounding effect in the gender disparity of CRC incidence outcome at the regional group level.

For the covariates adjustment and effect analysis, the ‘*’multiple linear regression’’* statistical test was employed. In this statistical test, the dependent variable remained the CRC incidence variable while the independent variables included covariates like ‘’wealth, Gini Index, education, health insurance access, and smoking’’, in addition to ‘’region’’ with its six independent groups. Multiple linear regression is appropriate when the dependent variable is continuous, with more than one independent variable measured at the nominal or scale levels (Daniel & Cross, 2013). The covariate analyses included in this statistical model were important to remove any potential confounding that may affect the validity of this study. The preceding is because covariates like income disparity and other socioeconomic variables are cited in the literature as potential risk factors for CRC incidence.

In interpreting the results of the statistical analyses in this study, the *p*-values and confidence interval estimates were used to determine statistical significance, which informed the decision to reject or fail to reject the stated null and research hypotheses. For

the spatial autocorrelation analysis, the Moran Index *(I*) was used. The value of the index was further converted to the standard *z*-value or score to test its statistical significance.

Threats to Validity

In this study, the main threat to external validity was the limitation of the sample to only the populations and settings covered by NSCR data set. This study's outcome may not be generalizable to other populations outside Nigeria (population validity) and settings (ecological validity), because the sampling frame and the ecological study design may not be easily adaptable to other populations and settings. Allen (2017) argued that external validity is the extent to which research findings can be generalized across other populations, groups, and settings.

The secondary data sets used in this study are vulnerable to internal validity and reliability challenges. Generally, the internal validity and reliability of a secondary data set in answering a RQ is a function of the data collection methods and accuracy. According to Clarke and Cossette (2000), the major threat to the validity and reliability of research results using secondary data sets is the precision and accuracy of data collection methods used in collecting the primary data. The foregoing is because the quality of a data set influences the quality of the resulting research results. Boo and Froelicher (2013) argued that the major threats to a secondary data set's validity and reliability arise from data collection methods, the definition of variables, missing data, measurement error, sample size, and sampling frame.

In this study, the NSCR data set was collected based on a mutually agreed upon CRC case definition and diagnosis. The quality of data from each cancer registry was

assessed using the indices of completeness, validity, and timeliness (Jedy-Agba et al., 2016). The methodology used in the data collection is also consistent with the global best practices in data collection for cancer registries. The above position is because the data set also included measuring the variables of interest, including the possible measurement errors, sample size and sampling frame, and the missing data (Jedy-Agba et al., 2016). However, the major threat to validity in using this data set was the fixed sample size, which predisposed the result of this secondary data analysis to type 11 error. Sub-group analyses may also have suffered some validity problems because of the resulting small sample sizes that may not be enough to statistically significantly power the sub-group analyses in the NSCR data set. Also, my poor understanding of the data entry process and not being involved in the sampling design and data collection may have been a limitation to the use of this data set (Clarke & Cossette, 2000). Another possible threat to validity was the use of hospital-based cancer registries (HBCR) to complement the population-based registries (PBCR), where there were no PBCR. The use of HBCR may have encouraged inflation or, otherwise, the under-reporting of CRC cases in some regions because hospital-based registries, unlike the population-based cancer registries, allow the movement of persons or CRC patients from one region to another, which will be a potential source of a validity challenge.

The ecological fallacy is a potential threat to this study's internal validity, as the outcome of this data cannot be used to make any inference at the individual level. The ecological fallacy will occur if the outcome of this study based on the group exposure data were held to be the same when the individual's exposure data are used. The ecological

fallacy can also be a threat to construct validity, because aggregate data often measure a different construct than their equivalent at the individual level (Schwartz, 1994). Ecological bias can also be explained by the heterogeneity of exposure and covariate levels within groups (Morgenstern, 1995).

The reliability and validity of the instrument used in determining the outcome variable of CRC cases in this data set were not disclosed, which may affect the validity of this study's outcome using the 2009-2013 NSCR data. Another source of validity problem was the possibility of missing data in the data set, which can pose a threat to validity, especially if the data were not ''Missing at Random'' (MAR) (Polit, 2010).

The operationalization of the exposure variables, especially the use of proxy pollutants and the assumption of cumulative exposures, are fraught with several misinterpretations, which constituted a threat to the construct validity because of measurement challenges.

Notwithstanding the potential threats to validity using this NSCR data set, the quality and integrity of the NSCR data set are reinforced as it is published in international peer-reviewed articles. The NSCR data was also used in calculating the national cancer estimate for Nigeria in the 2012 GLOBACAN cancer survey. Besides, the NSCR data set is a nationally representative data set. Therefore, threats to external validity when the results are applied in other populations was minimized.

To minimize the possibility of any threats to internal validity, I thoroughly reviewed the purpose and the summary reports, the codebooks, the manual of operations, and the previously published papers related to the data set, as recommended by Doolan and

Froelicher (2009). Also, before carrying out the statistical analysis using these data set, I examined the patterns of missing data to include them in the statistical analytical plan strategies to deal with missing data challenges (Boo & Froelicher, 2013). In the main, all the potential limitations of the included data sets were kept in mind when I was analyzing the data to inform appropriate valid conclusions.

Ethical Procedures

The NSCR data set for this study was only assessed after Walden University IRB approval (Approval number: 07-02-21-0602536) and approval by the IRB of the Nigerian Institute of Medical Research. The data set was accessed by completing the online data-sharing document and submitting it to the NSCR website. The data collection also respected the local law in Nigeria, governing the use of secondary data sets. Although this ecological study design did not lend itself to data collection from the individuals, their privacy and confidentiality were protected where such information was available as part of the NSCR data set. The preceding was to ensure obedience to the doctrine of beneficence that forbids harming research participants by respecting their autonomy and guaranteeing no harm to them while providing research benefits. Where and when necessary, further individual consent would have been obtained because consent is not a ‘’once-and-for-all event’’ but is subject to renegotiation over time (Statement of Ethical Practice for the British Sociological Association, 2004).

For emphasis, whether or not the individuals’ information were needed for this analysis, there is a need to protect their privacy by making sure that the data set is securely saved away from any unauthorized third party. According to Tripathy (2013), concerns

about the secondary use of data mostly revolve around potential harm to individual subjects and the issue of return for consent. However, as my research questions did not suggest the use of individual data, I only confirmed to the IRBs that the data is anonymous (Rudestam & Newton, 2015). As this data set is only for research and statistical analysis, I will store the data securely on my laptop and protect it with a password to ensure no access to unauthorized third parties (Centers for Disease Control and Prevention, 2018a). The data set will be destroyed immediately after use. No conflict of interest is applicable in this study.

Summary

In this chapter, I have briefly described the study purpose; and population, the six regions, as in figure 1, the methods of data collection, the secondary data set, the instrumentation and the operationalization of variables, the data analysis plan, the outcome and exposure measures, and how they will be used to answer the RQs. I have also described the sampling frame, sampling method, and strategy. The statistical analysis plan, including data cleaning and the proposed statistical tests in this study, were also discussed. The potential for confounding due to spatial autocorrelation, which is the tendency for adjacent regions to share similar outcome variables is discussed to determine the possibility of any clustering, dispersion, or randomness of the cases across the regions, given the contiguous nature of the states between regions. I have also discussed the potential threats to external and internal validity, including construct validity due to the ecological fallacy. How to address validity challenges in this study was also discussed in the statistical analysis plan. Ethical issues related to the use of the NSCR secondary data set, including the pledge to

confidentiality and privacy of the information on participants and the IRB process, in obedience to the beneficence doctrine and other research ethics, are discussed. The study results are presented in chapter four of this dissertation document.

Chapter 4: Results

Introduction

The main purpose of this study was to investigate whether living in the oil-polluted region in Nigeria is associated with excess CRC development. I also investigated spatial autocorrelation of CRC incidence rates across Nigerian regions, especially the contiguous regions. CRC gender disparity in the population under study was also examined by controlling for some socioeconomic factors like Gini Index, wealth index, education, access to health insurance, and smoking, to assess possible occupational hazards and to inform appropriate public health interventions to minimize disparity in the health outcome of interest. The RQs and the corresponding hypotheses were as follows:

RQ1: Is there a statistically significant relationship between living in the oil-polluted environment or region and CRC incidence in Nigeria?

$H_0$1: There is no statistically significant relationship between living in the oil-polluted environment or region and CRC incidence in Nigeria.

H_A1*:* There is a statistically significant relationship between living in the oil-polluted environment or region and CRC incidence in Nigeria.

RQ2: Is there a statistically significant relationship between the contiguity of the six regions in Nigeria and the spatial autocorrelation of CRC incidence?

$H_0$2: There is no statistically significant relationship between the contiguity of the six regions in Nigeria and the spatial autocorrelation of CRC incidence.

H_A2: There is a statistically significant relationship between the contiguity of the six regions in Nigeria and the spatial positive autocorrelation of CRC incidence.

RQ3: Is there a statistically significant relationship between gender and CRC incidence within and between Nigeria's various regions after adjusting for covariates, such as education, wealth, Gini Index, and access to health insurance?

$H_0$3: There is no statistically significant relationship between gender and CRC incidence within and between Nigeria's various regions after adjusting for covariates, such as education, wealth, Gini Index, and access to health insurance.

H_A3: There is a statistically significant relationship between gender and CRC incidence within and between Nigeria's various regions after adjusting for covariates, such as education, wealth, Gini Index, and access to health insurance.

In this chapter, I present the SPSS data analysis output, the descriptive statistics for the included variables, and the inferential statistics with the results of the analysis of the RQs. I begin the chapter by describing the data collection. This chapter concludes with a summary and transition to Chapter 5.

Data Collection

I obtained data on the crude and the ASR of the disease outcome of interest (CRC) from the NSCR covering 2009-2016. Only data from the 18 registries that responded to the data submission questionnaire from the NSCR were included in the sampling frame for analysis. The NSCR usually checks the data collected from the participating cancer registries for consistency and completeness and other data quality parameters, using CanReg4 & 5 software, depending on the version available in each of the registries that submitted data (Jedy-Agba et al., 2016)). Case definition and diagnosis were mutually agreed upon, while the ASRs were calculated for each of the population-based cancer

registries. I extracted only data from population-based cancer registries from each region that met the sampling inclusion criteria for this study.

I collected the exposure variable (PM 2.5) data online in real-time on the same date across the six regions from IQAir global air quality monitoring ground-based stations. IQAir uses data on PM 2.5 as a representative measure of environmental pollution (Viatte et al., 202). Data were collected and recorded from all the stations that represented the different local council areas of each state that make up each region. The mean values were calculated as PM 2.5 exposure values for each region and were entered into the data set. I used the U.S. air quality index (US AQI) as a unit of measurement. I also extracted data on covariates from the Nigerian 2018 NDHS. The NDHS used a two-stage stratified sampling technique that identified 74 sampling strata and 1,400 enumeration areas. All the data on covariates were available on the regional levels and were operationalized as continuous variables for the statistical analysis. I merged all three data sets (NSCR, NDHS, and IQAir) extracted on July 3, 2021, and checked the combined data for completeness before exportation to SPSS software for the respective statistical analyses.

Results

I exported the merged data set into the SPSS 27 software and downloaded the data to calculate both the descriptive and inferential statistics for the dependent, independent, and covariate variables to answer the RQs. The results are presented in Tables 4-28.

Descriptive Statistics

Table 4

Descriptive Statistics for the Continuous Variables in this Analysis

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
ASR_CRC	12	1.200	7.600	3.908	1.955	.478	.637	-.390	1.232
PM 2.5_Scale	12	46.000	76.000	61.333	11.292	-.364	.637	1.404	1.232
Gini Index	12	.130	.330	.232	.085	.107	.637	2.035	1.232
Wealth Index	12	5.400	48.300	22.450	16.319	.413	.637	1.225	1.232
Highest Level of Education	12	2.700	14.300	7.950	4.173	.162	.637	1.163	1.232
Access to Health Insurance	12	1.450	4.450	2.833	.968	.419	.637	-.207	1.232
Tobacco Smoking	12	1.750	5.200	3.233	1.324	.491	.637	1.478	1.232
Valid N (listwise)	12								

Table 5

Descriptive Statistics for Region Variable

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	northcentral	2	16.7	16.7	16.7
	northeast	2	16.7	16.7	33.3
	northwest	2	16.7	16.7	50.0
	southeast	2	16.7	16.7	66.7
	southsouth	2	16.7	16.7	83.3
	southwest	2	16.7	16.7	100.0
	Total	12	100.0	100.0	

Note. For each region, only one data point each was included for the male and female populations

Table 6

Descriptive Statistics for PM 2.5 Exposure Variable

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	46.00	2	16.7	16.7	16.7
	48.00	2	16.7	16.7	33.3
	64.00	2	16.7	16.7	50.0
	66.00	2	16.7	16.7	66.7
	68.00	2	16.7	16.7	83.3
	76.00	2	16.7	16.7	100.0
	Total	12	100.0	100.0	

Note. PM 2.5 exposures were measured using the US air quality index (US AQI).

Table 7

Descriptive Statistics for Gender Variable

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	6	50.0	50.0	50.0
	Female	6	50.0	50.0	100.0
	Total	12	100.0	100.0	

I ran the descriptive statistics tests for all the continuous variables using the SPSS 27 to calculate the measures of central tendency and dispersion. Tests for any violation of the assumptions of the statistical tests were also carried out. I also ran a separate descriptive statistic for all the included categorical variables to determine frequencies and

data distribution. I classified the gender variable as male = 1 and female = 2. The variable 'Region' was classified as 1 = Northcentral, 2 = Northeast, 3 = Northwest, 4 = Southeast, 5 = South South, and 6 = Southwest. PM 2.5 was operationalized as both categorical and continuous variables. I used the continuous variable data to test for the correlation between ASR_CRC and the PM 2.5 exposure variable. All the covariates were operationalized as continuous variables in the data set used for the statistical analysis.

The ASR of colorectal cancer incidence (ASR_CRC) had a mean value of 3.908, a standard deviation of 1.955, a minimum value of 1.200, and a maximum value of 7.600, as shown in Table 4. The continuous variable of PM 2.5 included in the data set had a mean value of 61.333, a standard deviation of 11.292, a minimum value of 46 and a maximum value of 76. The Gini Index covariate had a mean of 0.232, a standard deviation of 0.085, a minimum value of 0.130, and a maximum value of 0.330. The smoking covariate had a mean value of 3.233, a standard deviation of 1.324, a minimum value of 1.750, and a maximum value of 5.200. The descriptive statistics for other covariates are shown in Table 4. The descriptive statistics for the categorical variables showed that male and female cases were equally represented ($N = 12$) for all regions. Male cases had a total age-adjusted CRC incidence of 3.9833 (50.959%), while for female cases, it was 3.6333 (49.041%), as shown in Tables 7 and 10. As this is an ecological study, it is important to note that *N* does not represent the number of individual cases, but it represents male and female group mean data entered in this data set.

I examined the continuous variables to determine if they met the assumptions of normality for the statistical tests. The skewness statistic for ASR_CRC, PM 2.5, Gini

Index, and smoking were 0.478, 0.364, 0.107, and 0.491, respectively. Their respective kurtosis statistics were -0.390, -1.404, -2.036, and -1.478. All the other covariates had similar skewness and kurtosis values, as shown in Table 4. Skewness less than 1 and kurtosis less than 3 are indications of normal distribution of data (Daniel and Cross, 2013). Therefore, I assumed that all the continuous variables in the data set were approximately normally distributed and that any missing data were missing completely at random.

Inferential Statistics

The inferential statistical tests conducted to answer the RQs included the *Pearson* correlation test, one-way *ANOVA*, one-way *ANCOVA*, independent sample *t*- tests, *multiple linear regression*, and the Moran autocorrelation Index (*I)*. To ensure that the statistical tests were conducted without any violation of their assumptions, the data were examined for the assumptions. The assumptions for the one-way ANOVA*, and* ANCOVA include continuous dependent variable and categorical independent variable with more than two levels or independent groups. The Pearson correlation test requires that both the dependent and independent variables are measured at the scale or ratio levels and a linear relationship. The *t* -test requires a continuous dependent variable and a dichotomous independent variable. The multiple linear regression statistical test requires a continuous dependent variable with independent variables that can be continuous or categorical. In addition to the specific assumptions above, all the inferential statistical tests require normal distribution of the dependent variable, independence of observations, linearity between the dependent and independent variables, absence of significant

outliers, homoscedasticity, and homogeneity of variance. (Laerd Statistics, 2018). The criterion for statistical significance was set at 0.05. The findings of the statistical analysis, which are organized by the respective RQs are reported below:

Analysis of Variance and Covariance Tests.

I first performed the Pearson correlation test using the SPSS 27 to determine if there was an overall correlation between the age-standardized CRC rate (ASR_CRC) and the particulate matter (PM 2.5) exposure variables. After examining the data for the assumptions of the statistical tests, the one-way ANOVA and ANCOVA were performed to determine whether living in an oil-polluted region is associated with excess CRC incidence and to adjust for the covariate. Included in this model were ASR_CRC as the dependent variable and ‘’region’’ as the independent variable. The independent variable was categorized as 1=northcentral, 2=northeast,3=northwest, 4=southeast, 5=southsouth, and 6=southwest. A post-hoc test was conducted to assess variations of CRC incidence across the six regions, especially in comparison with the southsouth region. Gender was included in the ANCOVA test as a covariate to test its effect on CRC disparity in the population. The results are shown below:

Table 8

Correlations Between ASR_CRC and PM 2.5 Continuous Variables

		ASR_CRC	PM 2.5_Scale
ASR_CRC	Pearson Correlation	1	.181
	Sig. (2-tailed)		.574
	N	12	12
PM 2.5_Scale	Pearson Correlation	.181	1
	Sig. (2-tailed)	.574	
	N	12	12

Table 9

Between Subjects Factors

		Value Label	N
Regions	1	northcentral	2
	2	northeast	2
	3	northwest	2
	4	southeast	2
	5	southsouth	2
	6	southwest	2

Table 10

Group Statistics for Gender Categorical Variable

	Gender	N	Mean	Std. Deviation	Std. Error Mean
ASR_CRC	Male	6	3.983	1.807	.738
	Female	6	3.833	2.264	.924

Table 11

Descriptive Statistics for the Regional Mean Values of ASR_CRC in the One-Way ANOVA Test

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
northcentral	2	6.450	1.626	1.150	-8.162	21.062	5.300	7.600
northeast	2	5.300	1.838	1.300	-11.218	21.818	4.000	6.600
northwest	2	1.850	.919	.650	-6.409	10.109	1.200	2.500

southeast	2	3.450	.071	.050	2.815	4.085	3.400	3.500
southsouth	2	4.550	.355	.250	1.373	7.727	4.300	4.800
southwest	2	1.850	.212	.150	-.056	3.756	1.700	2.000
Total	12	3.908	1.955	.564	2.666	5.150	1.200	7.600

Table 12

Overall Significance and Model-Fit Test in the One-Way ANOVA Test

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	34.984	5	6.997	5.959	.025
Within Groups	7.045	6	1.174		
Total	42.029	11			

Figure 5

Plots Showing Variations of ASR_CRC Across the Six Regions

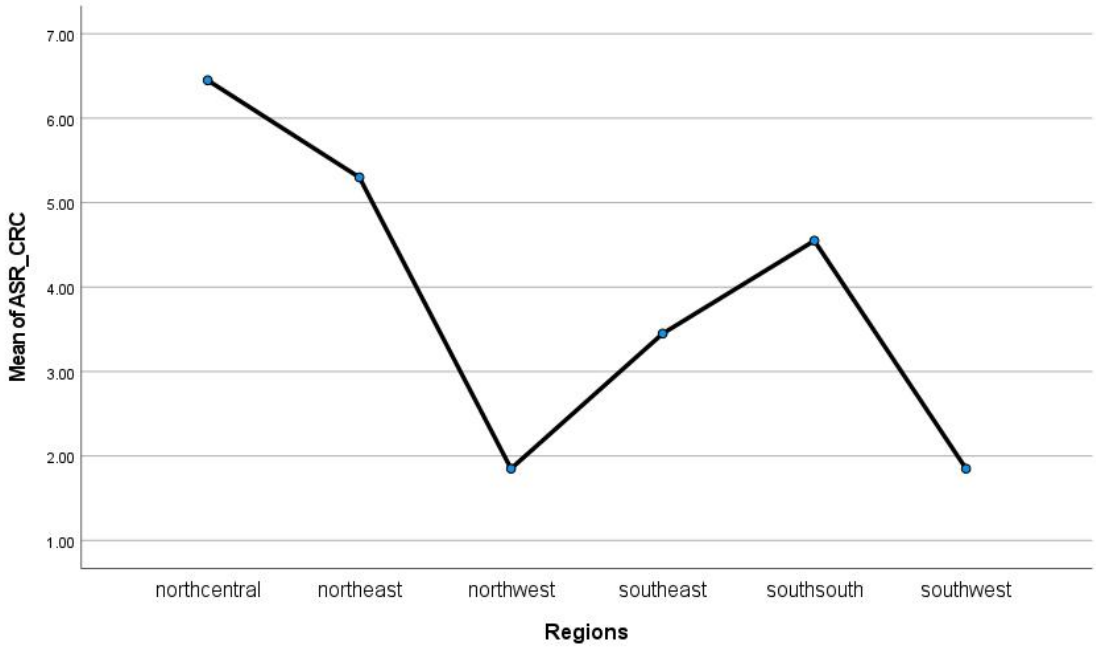


Table 13

Effect Size in the One-Way ANOVA Test

			95% Confidence Interval	
		Point Estimate	Lower	Upper
ASR_CRC	Eta-squared	.832	.000	.865
	Epsilon-squared	.693	-.833	.752
	Omega-squared Fixed-effect	.674	-.714	.735
	Omega-squared Random-effect	.292	-.091	.357

Table 14

Post-Hoc Multiple Comparison Results in the One-Way ANOVA Test

(I) Regions	(J) Regions		Std. Error	Sig.	95% Confidence Interval

		Mean Difference (I-J)			Lower Bound	Upper Bound
northcentral	northeast	1.15000	1.08359	.881	-3.1625	5.4625
	northwest	4.60000*	1.08359	.038	.2875	8.9125
	southeast	3.00000	1.08359	.190	-1.3125	7.3125
	southsouth	1.90000	1.08359	.549	-2.4125	6.2125
	southwest	4.60000*	1.08359	.038	.2875	8.9125
northeast	northcentral	-1.15000	1.08359	.881	-5.4625	3.1625
	northwest	3.45000	1.08359	.119	-.8625	7.7625
	southeast	1.85000	1.08359	.572	-2.4625	6.1625
	southsouth	.75000	1.08359	.976	-3.5625	5.0625
	southwest	3.45000	1.08359	.119	-.8625	7.7625
northwest	northcentral	-4.60000*	1.08359	.038	-8.9125	-.2875
	northeast	-3.45000	1.08359	.119	-7.7625	.8625
	southeast	-1.60000	1.08359	.689	-5.9125	2.7125
	southsouth	-2.70000	1.08359	.258	-7.0125	1.6125
	southwest	.00000	1.08359	1.000	-4.3125	4.3125
southeast	northcentral	-3.00000	1.08359	.190	-7.3125	1.3125
	northeast	-1.85000	1.08359	.572	-6.1625	2.4625
	northwest	1.60000	1.08359	.689	-2.7125	5.9125
	southsouth	-1.10000	1.08359	.897	-5.4125	3.2125
	southwest	1.60000	1.08359	.689	-2.7125	5.9125
southsouth	northcentral	-1.90000	1.08359	.549	-6.2125	2.4125
	northeast	-.75000	1.08359	.976	-5.0625	3.5625
	northwest	2.70000	1.08359	.258	-1.6125	7.0125
	southeast	1.10000	1.08359	.897	-3.2125	5.4125
	southwest	2.70000	1.08359	.258	-1.6125	7.0125
southwest	northcentral	-4.60000*	1.08359	.038	-8.9125	-.2875
	northeast	-3.45000	1.08359	.119	-7.7625	.8625
	northwest	.00000	1.08359	1.000	-4.3125	4.3125
	southeast	-1.60000	1.08359	.689	-5.9125	2.7125
	southsouth	-2.70000	1.08359	.258	-7.0125	1.6125

Table 15

Mean Estimates of ASR_CRC after Adjustment for Gender in the One-Way ANCOVA Test

			95% Confidence Interval	
Regions	Mean	Std. Error	Lower Bound	Upper Bound
northcentral	6.450	.835	4.303	8.597
northeast	5.300	.835	3.153	7.447
northwest	1.850	.835	-.297	3.997
southeast	3.450	.835	1.303	5.597
southsouth	4.550	.835	2.403	6.697
southwest	1.850	.835	-.297	3.997

Table 16

Pairwise Comparison of Mean ASR_CRC After Adjustment for Gender Covariate

(I) Regions	(J) Regions	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
northcentral	northeast	1.150	1.181	.375	-1.887	4.187
	northwest	4.600*	1.181	.011	1.563	7.637
	southeast	3.000	1.181	.052	-.037	6.037
	southsouth	1.900	1.181	.169	-1.137	4.937
	southwest	4.600*	1.181	.011	1.563	7.637
northeast	northcentral	-1.150	1.181	.375	-4.187	1.887
	northwest	3.450*	1.181	.033	.413	6.487
	southeast	1.850	1.181	.178	-1.187	4.887
	southsouth	.750	1.181	.553	-2.287	3.787
	southwest	3.450*	1.181	.033	.413	6.487
northwest	northcentral	-4.600*	1.181	.011	-7.637	-1.563
	northeast	-3.450*	1.181	.033	-6.487	-.413
	southeast	-1.600	1.181	.234	-4.637	1.437
	southsouth	-2.700	1.181	.071	-5.737	.337
	southwest	-3.331E-15	1.181	1.000	-3.037	3.037
southeast	northcentral	-3.000	1.181	.052	-6.037	.037
	northeast	-1.850	1.181	.178	-4.887	1.187
	northwest	1.600	1.181	.234	-1.437	4.637
	southsouth	-1.100	1.181	.395	-4.137	1.937
	southwest	1.600	1.181	.234	-1.437	4.637
southsouth	northcentral	-1.900	1.181	.169	-4.937	1.137
	northeast	-.750	1.181	.553	-3.787	2.287
	northwest	2.700	1.181	.071	-.337	5.737
	southeast	1.100	1.181	.395	-1.937	4.137
	southwest	2.700	1.181	.071	-.337	5.737
southwest	northcentral	-4.600*	1.181	.011	-7.637	-1.563
	northeast	-3.450*	1.181	.033	-6.487	-.413
	northwest	3.331E-15	1.181	1.000	-3.037	3.037
	southeast	-1.600	1.181	.234	-4.637	1.437
	southsouth	-2.700	1.181	.071	-5.737	.337

Research Question 1

RQ1: Is there a statistically significant relationship between living in the oil-polluted environment or region and colorectal cancer incidence in Nigeria? The *Pearson* correlation test showed a negligible positive correlation between the ASR_CRC variable and the PM 2.5 variable ($r = 0.181$), which was not statistically significant ($p=0.574$), as shown in Table 8. Regions and their value label are shown in Table 9. The one-way *ANOVA* test results showed that the overall difference in the mean value of ASR_CRC across the regions was statistically significant, F (5.6) =5.959, $p=0.025$), as shown in Table 12. The above evidence indicated that the model was a good fit for the data provided. The data also confirmed the presence of CRC incidence disparity in the Nigeria Population. However, cross-comparison of the CRC means across regions showed that the mean differences between regions were only statistically significantly different between northcentral and northwest regions ($P=0.038$) and between northcentral and southwest ($p=0.038$). No other comparison of means between regions was statistically significant. Specifically, the southsouth region that represents the region in an oil-polluted environment did not show any statistically significant mean difference in CRC incidence with other regions ($p=0.549$ with northcentral, $p= 0.976$ with the northeast, $p= 0.258$ with northwest, $p=0.897$ with the southeast, and $p=0.258$ with southwest), as shown in Table 14. After adjusting for gender as the covariate, the differences in the mean CRC values between the southsouth region and other regions were also not statistically significant ($p=0.165$ for northcentral, $p=0.553$ for the northeast, $p=0.071$ for northwest, $p=0.397$ for the southeast, $p=0.071$ for southwest), as shown in Table 16. In

addition, the ’estimates’ after adjustment for the covariate, as shown in Table 15, did not change when compared with the values in the descriptive statistics before adjustment, as shown in Table 11. However, in both the unadjusted and adjusted ASR_CRC mean values, the northcentral and northeast regions had higher mean ASR_CRC values compared with the oil-producing southsouth region, as shown in Table 11 and Figure 5. In this one-way ANOVA analysis, the region as a predictor variable accounted for about 83.20% (Eta-squared) of the variance in the CRC outcome variable, as shown in Table 13. Therefore, from the evidence provided after data analysis, I failed to reject the null hypothesis that there is no statistically significant difference in mean CRC incidence rate between living in the oil-producing southsouth region of Nigeria and the other regions, even after adjusting for the covariate. The foregoing evidence is supported by the lack of statistically significant correlation between PM 2.5 exposures and CRC outcomes in Table 8.

Spatial Autocorrelation Test

A spatial autocorrelation test was performed to determine if there was spatial dependence in the CRC outcomes across the regions, especially the contiguous regions. The crude incidence rate data was used and calculated based on the weighted average of male and female crude incidence rates. A binary spatial ‘’Rook’’ weighting matrix or contiguity table was set up to calculate the spatial weights. Contiguous regions are allocated the weight of ‘1,’ while non-neighbors are allocated a ‘0’ weight. The cross-product terms for all the non-zero weights and the total spatial weights were calculated and substituted in the Moran Index equation. The calculated Moran index was tested for

its statistical significance using the standardized Z-score and *p*-value. The results are shown below:

Table 17

Crude Incidence Rates of Colorectal Cancer

SNO	Region	Crude Incidence Rate
1	NC	2.40
2	NE	3.62
3	NW	1.12
4	SE	2.28
5	SS	2.60
6	SW	1.66

Note. NC=Northcentral, NE=Northeast, NW=Northwest, SE=Southeast, SS=Southsouth, SW=Southwest.

Table 18

Table of Deviations

Region	x	d= (x- x̄)	(x- x̄)2
NC	2.40	0.12	0.014
NE	3.62	1.32	1.742
NW	1.12	-1.16	1.346
SE	2.28	0	0
SS	2.60	0.32	0.102

SW	1.66	-0.62	0.384

Table 19

Binary Rook Connectivity Structure or Spatial Weight Matrix Derived from Figure 1

	NC	NE	NW	SE	SS	SW
NC	0	1	1	1	1	1
NE	1	0	1	0	0	0
NW	1	1	0	0	0	0
SE	1	0	0	0	1	0
SS	1	0	0	1	0	1
SW	1	0	0	0	1	0

Total spatial Weights ($\Sigma\Sigma Wij$) =16.
Note. Adjacent regions are allocated a weight of ‘’1,’’ while non-contiguous or non-adjacent regions are allocated ‘’0.’’

Table 20

Calculating the Cross-Product Terms

	NC	NE	NW	SE	SS	SW
NC	0	d1d2	d1d3	d1d4	d1d5	d1d6
NE	d2d1	0	d1d3	0	0	0
NW	d3d1	d3d2	d2d3	0	0	0
SE	d4d1	0	0	0	d4d5	0

SS	d5d1	0	0	d5d4	0	d5d6
SW	d6d1	0	0	0	6d5	0

*Note***.** d=deviations from the mean value

Table 21

Cross-Product Terms with Imputed Values

	NC	NE	NW	SE	SS	SW
NC	0	0.1584	-0.1392	0	0.0384	-0.0744
NE	0.1584	0	-1.530	0	0	0
NW	-0.1392	-1.530	0	0	0	0
SE	0	0	0	0	0	0
SS	0.0384	0	0	0	0	-0.1984
SW	-0.0744	0	0	0	-0.1984	0

Total Cross-product terms ($\sum\sum(x_i - \bar{x})(x_j - \bar{x})$ = -3.490

Moran Index $I = \frac{N}{W} \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$ (1).

Where N= the [illegible] idence; ΣΣwij=16; cross product terms $(x_i - \bar{x})(x_j - \bar{x})$ $\sum_i (x_i - \bar{x})^2$ and =

3.588

Substituting the values above in equation (1):

The Moran *I*= (6/16) x (-3.490/3.588) = -0.365.

The value of *I* shows a weak negative spatial autocorrelation

The ‘’Expected ‘value (E) of Moran’s *I* under the null hypothesis is:

E (*I*) = -1/N-1 (2)

E (*I*) = -1/(6-1) = -1/5 = -0.20

To test for the statistical significance of the null hypothesis, the Moran *I* is converted to the standard Z-score as follows:

Z= *1* -E(*I*)/√Var (*I*). (3)

The variance of Moran *I (*Va*r(I),* assuming normal approximation, is calculated as follows:

$$Var_N(I) = \left(\frac{1}{S_0^2 (n2-1)} (n^2 S_1 - nS_2 + 3S_0^2) \right) - E_N(I)^2 \quad (4)$$

Where:

$S_0 = \sum_{i=1}^{i=n} \sum_{j=1}^{j=n} W_{ij}$ The sum of the Spatial Weight Matrix $= 16$

$S_1 = \frac{\sum_{i=1}^{i=n} \sum_{j=i}^{j=n} (W_{ij} + W_{ji})^2}{2}$ If weight matrix symmetric then $S_1 = 2 \sum_{i=1}^{i=n} \sum_{i=1}^{j=n} W_{ij} = 32$

$S_2 = \sum_{i=1}^{i=n} (W_{i\bullet} + W_{\bullet i})^2$ The sum of the (i^{th}column + i^{th}Row)2 of weight matrix. If symmetrix $S_2 = 4 \sum_{i=1}^{i=n} W_{i\bullet}^2$

Substituting the values in equation 4 above:

Var(*I*) = -1.540

Z in equation 3 above = -0.018. P > 0.05

Therefore, the Moran *I* is negative, weak, and not statistically significant.

Research Question 2

RQ2: Is there a statistically significant relationship between the contiguity of the six regions in Nigeria and the spatial autocorrelation of CRC incidence? The Moran Index (*I)* was calculated to be -0.350 based on equation (1) and the values from Tables 17-21 above, which shows a weak negative spatial autocorrelation. From the *Z*-value of Moran *I*, the *P*-value was found to be >0.05, which showed that the negative autocorrelation was not statistically significant at the 0.05 statistical significance criterion. Therefore, I failed to reject the null hypothesis of no statistically significant autocorrelation. The Moran *I* value ranges from + 1 to -1. Values approaching +1 or -1 are said to be perfectly positively or negatively corrected. +1 value is evidence for clustering of like values while -1 value indicates dispersion or clustering of unlike values. Values approaching zero indicate randomness of cases or independence of observations of cases. From the statistical evidence provided in the results, I failed to reject the null hypothesis and concluded that there was no statistically significant autocorrelation of CRC across the regions investigated in this study.

Independent Sample t- Tests and Multiple Linear Regression Tests

The independent sample *t*-test was conducted using SPSS 27 to test the effect of gender on the CRC disparity across the six regions. After data cleaning, gender was entered as the independent variable, while the dependent variable was the age-standardized CRC incidence rate (ASR_CRC). Gender was dichotomized as 1=male and 2=female. To test for the impact of covariates, a multiple linear regression test was conducted after checking the data for statistical test assumptions using some regression

diagnostics. In model 1, gender was entered alone as the predictor or independent variable. In model 2, adjustments for all covariates were made by including Gini Index, Wealth Index, education, access to health insurance, and smoking in the model. The gender variable was also included in model 2. All covariates were operationalized as continuous variables. The results are shown below:

Table 22

The Independent Sample-Test Results

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
ASR_CRC	Equal variances assumed	.101	.757	.127	10	.902	.15000	1.18268	-2.48517	2.78517
	Equal variances not assumed			.127	9.531	.902	.15000	1.18268	-2.50284	2.80284

Table 23

Group Statistics for Gender Categorical Variable

	Gender	N	Mean	Std. Deviation	Std. Error Mean
ASR_CRC	Male	6	3.9833	1.80712	.73775
	Female	6	3.8333	2.26421	.92436

Table 24

Effect Size on the Multiple Linear Regression Test

				95% Confidence Interval	
		Standardizer[a]	Point Estimate	Lower	Upper
ASR_CRC	Cohen's d	2.04845	.073	-1.061	1.203
	Hedges' correction	2.21996	.068	-.979	1.110
	Glass's delta	2.26421	.066	-1.069	1.195

Note: Cohen's d uses the pooled standard deviation.

Table 25

Correlation Comparison for all Variables in the Multiple Linear Regression Test

		ASR_CRC	Gender	Gini Index	Wealth Index	Highest Level of Education	Access to Health Insurance	Tobacco Smoking
Pearson Correlation	ASR_CRC	1.000	-.040	.243	-.340	-.216	-.295	.156
	Gender	-.040	1.000	.000	.000	.000	.000	.000
	Gini Index	.243	.000	1.000	-.969	-.967	-.605	-.650
	Wealth Index	-.340	.000	-.969	1.000	.987	.707	.454
	Highest Level of Education	-.216	.000	-.967	.987	1.000	.730	.477
	Access to Health Insurance	-.295	.000	-.605	.707	.730	1.000	.047
	Tobacco Smoking	.156	.000	-.650	.454	.477	.047	1.000
Sig. (1-tailed)	ASR_CRC	.	.451	.223	.140	.251	.176	.314
	Gender	.451	.	.500	.500	.500	.500	.500
	Gini Index	.223	.500	.	.000	.000	.019	.011
	Wealth Index	.140	.500	.000	.	.000	.005	.069

Highest Level of Education	.251	.500	.000	.000	.	.003	.058
Access to Health Insurance	.176	.500	.019	.005	.003	.	.442
Tobacco Smoking	.314	.500	.011	.069	.058	.442	

.

Table 26

Model Summary for the Multiple Linear Regression Test

				Std.	Change Statistics					
Model	R	R Square	Adjusted R Square	Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
1	.040[a]	.002	-.098	2.04845	.002	.016	1	10	.902	
2	.913[b]	.834	.635	1.18131	.832	5.014	5	5	.051	1.695

Table 27

Overall Significance and Model-Fit Test for the Adjusted Multiple Linear Regression Test

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.067	1	.067	.016	.902[b]
	Residual	41.962	10	4.196		
	Total	42.029	11			
2	Regression	35.052	6	5.842	4.186	.069[c]
	Residual	6.977	5	1.395		
	Total	42.029	11			

Table 28

Coefficients of the Multiple Linear Regression Test

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	4.133	1.870		2.210	.052		
	Gender	-.150	1.183	-.040	-.127	.902	1.000	1.000
2	(Constant)	-24.395	23.473		-1.039	.346		
	Gender	-.150	.682	-.040	-.220	.835	1.000	1.000
	Gini Index	65.374	60.637	2.855	1.078	.330	.005	211.222
	Wealth Index	-.394	.267	-3.285	-1.473	.201	.007	149.785
	Highest Level of Education	2.726	.650	5.820	4.197	.009	.017	57.906
	Access to Health Insurance	-1.075	.670	-.532	-1.605	.170	.302	3.313
	Tobacco Smoking	1.110	1.028	.752	1.080	.330	.068	14.610

Note: Dependent Variable: ASR_CRC

Research Question 3

RQ3: Is there a statistically significant relationship between gender and CRC incidence within and between Nigeria's various regions after adjusting for covariates, such as education, wealth, Gini Index, access to health insurance, and smoking? From the statistical analysis results, gender did not statistically significantly predict CRC outcomes across the regions in the model, t (10) =0.127, p=0.902, as shown in Table 22. The males had a mean age-standardized CRC value of 3.9833 (50.959%), while the

females had a value of 3.8333 (49.041%), as shown in Table 23. The 95% confidence interval value shown in Table 22 shows that the difference between the mean values for males and females is not statistically significant as the interval contains the ‘’zero’’ value. The small 1.918 % difference in the mean values of the outcome variable between males and females may be responsible for the lack of statistical significance. The effect size measured by Cohen’s distance based on the pooled standard deviation was only7.3%, as shown in Table 24. The above confirmed that gender was only able to explain 7.3% of the variation in the outcome or dependent variable (ASR_CRC). After adjusting for covariates, there was no statistically significant correlation between ASR_CRC and gender (p=0.451). Also, none of the covariates had a statistically significant correlation with ASR_CRC, as shown in Table 25. The multiple linear regression model summary indicated that the effect size or coefficient of determination, represented by R squared was only 0.002. The change between model 0 and model 1 was only 0.002 (0.20%). However, model 2, that included covariates, showed an effect size or coefficient of determination value, represented by R squared, of 0.913, with an increase in value from model 1 of 0.837 (83.70%), as shown in Table 26. The *ANOVA* table showed that both models were not a good fit for the data provided, (F (1,10) 0.016, p=0.902 for model 1 and F (6,5) 4.196, p= 0.069 for model 2), as shown in Table 27. In model 1 that included only gender as the predictor variable, the partial regression coefficient (B) for gender was not statistically significant even after adjusting for covariates in model 2, as shown in Table 28.

For the effect of covariates on the dependent variable, as shown in Table 28, the Gini Index had the greatest effect on the ASR_CRC dependent variable. A unit increase in the Gini Index, that represents inequality in the population, increased the CRC incidence rate 65 times, controlling for other variables. For the wealth index that represents income level for the family households, a unit increase resulted in a decrease in ASR_CRC incidence rate by 0.394. Also, a unit increase in access to health insurance decreased the CRC incident rate in the population by about 1.08 times. However, a unit increase in higher education increased the incidence rate of CRC by 2.726, and a unit increase in tobacco smoking increased the incident cases by 1.10 times. None of these effects by covariates were statistically significant, except for the highest education variable (p=0.009), as shown in Table 28. The variance of the estimated coefficients, otherwise known as the variance inflation factors (VIF), of Gini Index, wealth index, and higher education were found to be high, showing that they were highly correlated with at least one of the variables included in this model, while smoking, access to health insurance, and gender did not show any multicollinearity with other included variables, as seen in Table 28. As a rule, any VIF value more than 10 is said to be inflated and a sign of multicollinearity. Put together, I failed to reject the null hypothesis that there was no statistically significant association between gender and the risk of CRC in the population under study (p=0.902), even after adjusting for covariates (p=0.069).

Summary

The findings suggested the presence of CRC incidence disparity in the studied populations. However, the study revealed no statistically significant association between

living in an oil-polluted region and the risk of excess colorectal cancer development in the studied Nigeria populations, even after adjusting for gender. The study findings also showed no statistically significant relationship between gender as a predictor variable and the incidence of colorectal cancer across the six regions. There was no statistically significant difference in incident colorectal cancer outcome between male and female populations. The adjustment for covariates did not change the non-statistically significant relationship between gender and incident colorectal cancer across the six regions. When the incident colorectal cancer outcome variable was regressed over the gender predictor variable, the effect size was only about 0.2%, which increased to 83.40% with the inclusion of covariates in the model. It was also important to note that there was no statistically significant correlation between incident colorectal cancer and the PM 2.5 exposure variable. The study findings also confirmed a weak negative spatial autocorrelation of colorectal cancer amongst the contiguous regions, which was not statistically significant ($P>0.05$). The interpretation of the main result findings, including the strengths, limitations, recommendation for future research, and the social change implications of this research in the studied population are discussed in chapter 5 of this dissertation document.

Chapter 5: Discussion, Conclusions, and Recommendations

Introduction

The primary purpose of this study was to investigate the association between living in an oil-polluted region in Nigeria and the risk of excess CRC using data from NSCR, IQIR, and NDHS. I also investigated the effect of gender on CRC disparities and the spatial autocorrelation of CRC across the regions. I conducted this study to determine whether living in the oil-producing South South region in Nigeria poses a greater risk of CRC incidence than living in other regions. In this study, data analysis included frequencies, measures of central tendency, and dispersion. Inferential statistical tests included the Pearson correlation test, one-way ANOVA and ANCOVA, t-test, multiple linear regression, and Moran spatial autocorrelation. The main findings were that the mean difference in age-standardized CRC incidence between the oil-producing South South region and other regions, including the male and female mean values, was not statistically significant, which did not change even after adjustment for the covariates. The study also found a weak negative autocorrelation, which was not statistically significant, suggesting no spatial dependence (clustering or dispersion) of CRC outcomes in the population.

Interpretation of the Findings

Within the methodological limitations of the ecological study approach that lacked information on individual exposure, this study found no evidence of an association between living in an oil-polluted environment and the risk of excess CRC. The literature reviewed showed an inconclusive association between oil pollution and CRC risk in

different populations. Yuan et al. (2018) and other studies confirmed a statistically significant association, while Vargas et al. (2020) confirmed no association.

The ecosocial theoretical framework that grounded this study was appropriate and consistent. This is because ecosocial epidemiological theory integrates the historical, socioenvironmental, and ecological perspectives of disease origin to inform disease distribution within a population (Krieger, 2011). According to Krieger (2011), ecosocial theory enabled the Faber and Harvard Cancer Center working group to identify gaps in knowledge about cancer inequality by addressing the social inequality at multiple levels. Canchola et al. (2017), Kichuchi et al. (2017), and Zhang et al. (2018) all implicitly used the ecosocial theoretical framework to explain the relationship between social-environmental factors, such as living in poorly built neighborhoods, stress, low socioeconomic status, and the risk of developing CRC. Based on this theory, I used the one-way ANOVA and ANCOVA to show the overall disparity of CRC outcomes in the populations based on their differential socioenvironmental factor exposures after controlling for covariates. The interpretation of the results of the research findings by RQ follows.

Research Question 1

The one-way ANOVA and ANCOVA statistical test results revealed a statistically significant difference in the mean value of CRC across all regions *(p* = 0.025), which was not statistically significant after adjusting for gender in the model (*p* = 0.051). However, cross-comparison of means across regions showed that the South South region that represents the people living in the oil-polluted environment did not show any statistically

significant difference in means of CRC incidence with other regions ($p = 0.549$ with Northwest, $p = 0.976$ with Northeast, $p = 0.258$ with Northwest, $p = 0.892$ with Southeast, and $p = 0.258$ with Southwest), even after adjusting for gender. Thus, although the study findings indicated overall CRC disparity in the population, the South South region did not have any statistically significant excess CRC outcomes compared to other regions.

Previous studies showed an inconclusive and varied relationship between living in an oil-polluted environment and the risk of excess CRC. Yuan et al. (2018) found a positive association between living within a 10 km radius of a petrochemical plant and the risk of CRC, which reinforced the earlier finding from Fano et al.'s (2006) ecological study that living in an oil-producing region was associated with CRC increase in a Taiwanese population. Lopez-Abente et al. (2012) used an ecological study design to confirm that living near a polluted industrial area increases the risk of CRC incidence in the Spanish population (RR: 1.26; 95% CI, 1.02-1.46). Ayuso-Alvarez et al. (2020) also confirmed in their study in the Spanish population that living near an industrial pollution site (5 km radius) increased the risk of CRC incidence (HR = 1.09, 95%CI, 1.05-1.15). Nnagu et al. (2016) confirmed an association between living in the oil-producing region and health problems in the Nigerian population, not necessarily CRC. However, Vargas et al. (2020) found no association between living in an oil-polluted environment and the risk of excess CRC in the Ecuadorian population, which is supported by an earlier study in Western Australia by El-Zaemays et al. (2017) that confirmed no association between CRC and 18 selected occupational agents.

This study's findings showed no correlation between PM 2.5 exposure and incident CRC in the sampled population. It is expected from the social problem of CRC disparity that those living in the oil-producing region should have a higher mean value of PM 2.5 exposure. However, the data collected for the PM 2.5 exposure variable, which served as a proxy pollutant for all potential environmental risk factors of CRC, was lowest in the South South region, compared to other regions, suggesting that living in the oil-producing region does not confer a higher level of exposure to the putative risk factors of CRC like PM 2.5. Previous researchers reported a varied relationship between PM 2.5 exposures and CRC outcomes. Jemwitheesuk et al. (2020), in research conducted in Thailand, demonstrated a statistically significant association between PM 2.5 exposure and CRC risk. Chu et al. (2020) provided temporal evidence through prospective studies in the United States that supported a positive association between CRC and PM 2.5 exposures. However, Wong et al. (2016) reported no association between PM 2.5 exposures and incidence of CRC in a cohort study in Hong Kong, while Turner et al. (2017) reported a positive association with PM 2.5 only in combination with NO2 exposures. This study added to the list of studies that do not support a statistically significant association between living in the oil-polluted region and the risk of excess CRC, including no evidence of a statistically significant correlation between PM 2.5 exposures and CRC incidence.

Research Question 2

The spatial autocorrelation analysis in this study showed a weak negative Moran spatial correlation index (*I*) of -0.350, which was not statistically significant (Z= -0.018, *p*

> 0.05). This finding suggests the randomness of CRC cases across the studied populations. A Moran Index *(I*) of +1 shows a perfect positive spatial autocorrelation (Mansori et al., 2018), which means clustering CRC cases of similar values like high and high values or low and low values. A Moran Index of -1 indicates a perfect negative autocorrelation of dissimilar values, which is an indication of dispersion or spreading of cases. Values close to 0 indicate randomness or independence of cases. Spatial autocorrelation in any disease outcome occurs when the value of a variable at a specified geographic location depends on its values at adjacent locations (Rosenberg, 2000). This is because nearer things are more related than distant things (Al-Ahmadi & Al-Zahrani, 2013).

There are extant pieces of evidence to show that different types of cancer exhibit different degrees of spatial autocorrelation at different cities and distances, including gender disparity (Al-Zahrani, 2013). Li et al. (2017) specifically conducted a spatio-temporal analysis of CRC incidence in Guangzhou, China. The result showed a global Moran Index of 0.527 (Z-11.06, $p < 0.001$), suggesting positive spatial autocorrelation and cluster distribution pattern of CRC in the population. Mansouri et al. (2018) used an ecological study design in Tehran to show positive autocorrelation of CRC outcomes. Also, Dechello and Sheeham (2007) confirmed a statistically significant autocorrelation of CRC in Massachusetts in the United States. However, the study in Kuala Lumpur conducted by Shah et al. (2014) showed no spatial autocorrelation of CRC in this population under investigation ($I = 0.007$). Therefore, my study finding on spatial autocorrelation of CRC corroborates the findings of some previous studies on the

existence of no spatial autocorrelation of CRC in different populations. This knowledge may be important in implementing an equitable public health intervention across the six regions. The above suggests that public health interventions should be based on the burden of CRC cases in each region rather than on contiguity to the oil-polluted region. The evidence of no statistically significant spatial autocorrelation also suggests that the statistical tests in this study did not violate the assumptions of independence of CRC observations (Daniel & Cross, 2013).

Research Question 3

The findings from the *t*-test statistical analysis showed that gender could not statistically significantly predict CRC outcomes across the six regions ($p = 0.902$), even after adjusting for the included covariates ($p = 0.069$). These findings suggest that being a male or female was not a statistically significant risk factor for CRC incidence in the population of interest. When the CRC incidence outcome variable was regressed on the gender predictor variable, the effect size was only 0.002, which increased to 0.832, after adjusting for covariates like Gini Index, wealth index, access to health insurance, education, and smoking, in the multilinear regression model. The partial regression coefficient (B) of gender (-150) in the first model, which included gender only as the predictor variable, was not different in the second model, which included gender and all the covariates, which showed that gender could not possibly confound the result findings in this ecological study.

Previous studies in different populations indicated gender disparity in CRC outcomes. The retrospective study conducted by Gao et al. (2008) in a Canadian

population showed that men were at higher risk of developing CRC than women within the same age cohort. Nguyen et al. (2009) conducted a systematic review of evidence and meta-analysis to provide quantitative pooled risk estimates that indicated a statistically significant association between gender and advanced colorectal neoplasia (RR = 1.83; 95% CI: 1.69-1.97), with men showing a higher risk than women. White et al. (2018) used cross-sectional national data on CRC incidence rates by age in the United Kingdom to conclude that men had a higher risk of developing CRC than women. The evidence of CRC gender disparity reported by the previous studies increased with age. However, the gender disparity with age narrowed over time, as women over 65 years were shown to be more at risk than men in the same age cohort (Kim et al., 2015). Overall, my study findings did not support previous studies showing evidence of gender disparity, with men being more at risk than women. The findings of no gender disparity in this study may inform gender equity in CRC screening across the six regions in Nigeria.

The multiple linear regression analysis revealed the various effects of the included socioeconomic covariates on CRC incidence in the studied population. As expected, based on previous studies, the Gini Index, which represents inequality in the population, had the highest effect or impact on CRC outcome. By holding other variables constant, a unit increase in inequality (Gini index) in the population increased the CRC incidence more than 65 times. Inequality manifests its effect through the pathways of poverty and stress. Kim et al. (2010) confirmed a positive association between living near a poor neighborhood defined by its proximity to industrial pollution and the risk of cancers. Kikuchi et al. (2017) confirmed an association between perceived stress and CRC, which

was reinforced by Kruk et al. (2019), whose research findings confirmed the positive relationship between psychological stress and chronic diseases like cancer. The regression analysis also revealed that a unit increase in the wealth index, which is a measure of household income, marginally decreased the incidence of CRC in the population. The above finding is consistent with the previous finding by Orlewska et al. (2018) in Poland, which confirmed a positive association between an increase in socioeconomic status, including social capital, and a decrease in CRC incidence. The statistical analysis results also suggested that an increase in access to health insurance will lead to a decrease in CRC outcomes. A unit increase in tobacco smoking resulted in an increase in CRC in the population, which supported previous studies that associated tobacco smoking with an increase in CRC outcomes (Chao et al.,2000; Hurley et al., 2013; Parajuli et al., 2014). However, acquiring higher education was surprisingly linked with an increase in CRC incidence in this study. The above evidence contradicted the previous study findings of Kim et al. (2010) of low CRC incidence with higher education status. However, a more recent study supported my study findings that acquiring higher education may be associated with an increase in CRC (Leufkens et al., 2012). From the descriptive statistics of the covariates, there was evidence to show that the northcentral and northeast regions with the highest incident CRC also showed high proportions of covariates-putative risk factors. The relatively lower incident CRC in the southsouth region may be explained in part by the low-income disparity, high wealth index, and high access to health insurance in this population, compared to other regions, except the southwest region. The findings of this study will assist public health interventions to

reduce CRC disparity by addressing the covariate risk factors and motivate social change in the population.

Limitations of the Study

The major strength of this study is the inclusion of spatial terms that exposed no statistically significant spatial dependence of the disease outcome in the population, which is important in addressing any clustering or spreading of CRC and the allocation of resources to minimize the disparity across the six regions. In addition, the quantitative ecological study design used in this study can be considered a strength because the study is primarily designed to provide data at the group or regional level to support the ongoing scholarly inquiry on the impact of oil pollution on health outcomes among communities living in Nigeria's oil-producing region. Also, the online real-time data collection of the PM 2.5 exposure variable is a strength, as it minimized potential measurement errors and increased the validity of the study outcome. Finally, worthy of note is the use of NSCR, which is also used in the calculation of national and international cancer estimates, with no report of any misleading or inaccurate data.

Despite the strengths identified in this study, the study is not without limitations. The methodological problem inherent in assuming individual association based on group data is a challenge, otherwise known as the ecological fallacy. Therefore, the major limitation of this study is the absence of individual data on CRC risks, PM 2.5 exposures, and socioeconomic variables. A causal relationship may, therefore, be difficult to establish (Setia, 2016). The ecological fallacy inherent in this study is a potential threat to this study's internal validity, as the outcome of this data cannot be used to make any valid

inference at the individual level. The ecological fallacy usually occurs in any study if the outcome of the study based on the group exposure data were held to be the same when the individual's exposure data are used. This ecological fallacy occurs easily when group-level data are used as proxies for the unavailable individual-level exposure. This ecological fallacy can also threaten construct validity because aggregate data often measure a different construct than their equivalent at the individual level (Schwartz, 1994). Therefore, this study's outcome cannot be used to make any informed inference of CRC disparity in the studied population at the individual level of exposure. The above limitation is common with most ecological study designs and can potentially confound studies. The use of only the population-based cancer registries might have limited the sample size. For instance, in some regions like the northwest with only one population-based cancer registry, the data collected may not be a true representation of the northwest region, which may result in underestimates of CRC rates in the region. Also, data collection of the CRC outcome in most registries covered 8 years, while a few others were about 5 years, which may affect the number and rate of CRC cases in favor of the regions that collected data over a longer period. Only six regions were used as ecological units, which may challenge the sample size. The above limitation can be overcome in future research by using the ''State or Local government areas'' as the ecological units of measurement. The use of proxy pollutants in this study, in my considered opinion, may not be a true representation of all the possible risk factors, which may pose an internal validity challenge. Besides, some of the covariates included in the multiple linear regression test showed multicollinearity with other variables, which may affect the

outcome of the test on the effect of covariates on the relationship between CRC incidence outcomes and gender.

Recommendations

This study did not find any statistically significant association between living in an oil-producing region and the risk of excess CRC incidence in Nigeria. However, this finding should be interpreted cautiously within the contexts of the strengths and limitations of this study. Given the evidence of the major limitations, I would recommend that further studies in this subject domain should use the local government area data, if available, as the ecological units, in place of regions. This will increase the sample size of the ecological unit of measurement and the validity of the research findings. The evidence of no statistically significant spatial autocorrelation was established in this study. However, the evidence was only on the global Moran autocorrelation index. Therefore, further research should also include the local Moran index, which is a local indicator of spatial autocorrelation (LISA) that provides details on the hot spots in areas within the regions that may show spatial autocorrelation to guide targeted and efficient public health interventions and practice. To minimize the influence of ecological fallacy, I recommend that future research on this knowledge domain should use the ‘’multilevel model,’’ which will include both the ecological and individual data to make informed decisions. Multilevel studies will increase both the internal and construct validity by providing associations or correlations between two variables at the individual and ecological unit levels (Schwartz, 1994). The multilevel construct or model will allow researchers to simultaneously make decisions on the variability of outcomes at the

individual and group levels and draw inferences regarding causes of variations (Diez Roux, 2000).

Implications

This study contributes to positive social change by providing evidence-based data on the relationship between environmental exposures and the risk of chronic disease like CRC. The evidence provided in this study may persuade the people living in the South South region to accept that living in the oil-producing region in Nigeria does not necessarily confer any higher risk of poor health outcomes than other regions. This will reduce the level of agitation in the South South region, especially the Niger Delta population, and may make the agitating youths a more positive social change agent (Chikwem, & Duru, 2018). Besides, addressing disparity in colorectal cancer across the six regions in Nigeria based on this study's evidence may support positive health behavior and other institutional reforms that will motivate social change across the country. According to de la Salonniere (2017), social change is about the transformation of behavior, culture, social institutions, and social structure over time. The use of spatial autocorrelation may contribute new knowledge in chronic disease epidemiology. In conclusion, the outcome of this study motivates social change and new knowledge in chronic disease epidemiology. This study's methodological and theoretical approaches are considered appropriate and cost-effective with a robust implication that may lead to the creation and advancement of health equity in the populations studied.

Conclusion

Given the weight of evidence provided in this study, it is reasonable to surmise that it found no association between living in an oil-polluted region in Nigeria and the risk of excess colorectal cancer incidence. The findings from this study also did not support any gender disparity and its implications of occupational hazards. The CRC incidence outcomes across the six regions occurred at random, as the weak negative autocorrelation was not statistically significant. In this study, I used the quantitative correlational ecological study design and secondary data sets to establish the possible association between living in an oil-polluted region in Nigeria and CRC development. The results of this study did not support the positive relationship between living in the southsouth oil-producing region in Nigeria and the excess risk of CRC, including gender disparity, after controlling for covariates at the regional level. The study also found a non-statistically significant negative autocorrelation of CRC incidence between contiguous regions, suggesting that the CRC outcomes occurred randomly with no spatial dependence. The non-inclusion of all possible putative socioenvironmental risk factors suggests the importance of life-course study approaches. Also of public health interest in this study's findings is the evidence that income disparity (Gini index), household wealth, access to health insurance are social factors that can influence health disparity in a population. The above evidence is consistent with the findings of the WHO committee on the social determinants of health. The foregoing evidence suggests that public health interventions towards reducing CRC disparity in Nigeria must address the social problems of income inequality and poverty, including access to quality and affordable

healthcare insurance in the studied population. It is also important to infer from the study findings that both the male and the female populations are at about the same risk of CRC incidence. The relatively low CRC incident rates in the southsouth and southwest regions may be explained by the possible moderating effect of covariate factors, such as low-income disparity, high household income, and access to health insurance, compared with other regions with higher CRC incident rates. The above indicates that living in an oil-polluted region may not be a single putative risk factor for CRC development. Hence, the need for a lifecourse approach to chronic disease epidemiology. In making an informed conclusion based on this study's outcome, especially at the individual level, there is a need to take into cognizance the potential confounding challenge of ecological fallacy. The ecological fallacy can be a threat to construct and internal validity because aggregate data often measure a different construct than their equivalent at the individual level (Schwartz, 1994). Therefore, to improve our knowledge in this association and chronic disease epidemiology, further studies should include a larger sample size of ecological units and a multilevel model study design to improve the internal and construct validity of the study outcome. Besides, the absence of temporality evidence and other causality criteria in the relationship between the variables of interest makes causality inference in this study difficult. In the main, this study found no association between living in the oil-polluted region in Nigeria and excess CRC development. However, this study's findings should be interpreted cautiously within the limitations identified in the study.

References

Abdulwahab, L., Dahalin, Z., & Galadima, M. D. (2011). Data summary and planning analysis of the determination of user acceptance of telecentre. *Journal of Information System: New Paradigms, 1*(1), 2224-2465. http://www.seaofsci.com/jisnp/Volume1_Issue1/Determinants%20of%20User%20Acceptance.pdf

Al-Ahmadi, K., & Al-Zahrani, A. (2013). Spatial autocorrelation of cancer incidence in Saudi Arabia. *International Journal of Environmental Research and Public Health*, *10*(12), 7207–7228. https://doi.org/10.3390/ijerph10127207

Aleksandrova, K., Pischon, T., Jenab, M., Bueno-de-Mesquita, H. B., Fedirko, V., Norat, T., Romaquera, D., Knuppel, S., Boutron-Ruaut, M. C., Dossus, L., Dartois, L., Kaaks, R., Li, K., Tionnedland, A., Overvad, K., Quiros, L. R., Buckland, G., Sanchez, M. J., Dorronsoro, M., ... Khaw, K. (2014). Combined impact of healthy lifestyle factors on colorectal cancer: A large European cohort study. *BMC Medicine, 12*, Article 168. https://doi.org/10.1186/s12916-014-0168-4

Allen, M. (2017). *The Sage encyclopedia of communication research methods* (Vols. 1-4). SAGE Publications. https://doi.org/10.4135/9781483381411

Althubaiti A. (2016). Information bias in health research: Definition, pitfalls, and adjustment methods. *Journal of Multidisciplinary Healthcare*, *9*, 211–217. https://doi.org/10.2147/JMDH.S104807

Ana, G. R. (2011). Air Pollution in the Niger Delta Area: Scope, Challenges and Remedies. http://www.intechopen.com/books/the-impact-of-airpollution-on-

health-economy-environment-and-agricultural-sources/air-pollution-in-the-niger-delta-area-scopechallenges-and-remedies

Antonius, R. (2013*). Interpreting quantitative data with IBM SPSS statistics* (2nd ed.). SAGE Publications. https://doi.org/10.4135/9781526435439

Ayuso-Alvarez, A., Garcia-Perez, J., Trivino, J., Gonzalez, T. M, Ramis, S. R., Lopez-Abente-G., Galan, M., & Fernandez-Navarro, P. (2020). Association between proximity to industrial chemical installations and cancer mortality in Spain. *Environmental Pollution, 360*, Article 113869. https://doi.org/10.1016/j.envpol.2019.113869

Badmos, K. B., Rotimi, O., Lawal, A. O., Osinowo, A. O., Habeebu, M. Y., & Abdulkareem, F. B. (2018). Histological histopathological report of colorectal carcinoma: A 5-year audit in Lagos. *Nigerian Postgraduate Medical Journal, 25*(1), 32-36. https://doi.org/ 10.4103/npmj.npmj_184_17

Basaleem, H. O., & Al-Sakkaf, K. (2004). Colorectal cancer among Yemeni patients: Characteristics and trends. *Saudi Medical Journal, 25*, 1002–1005. https://www.researchgate.net/publication/8388621

Ben-Shlomo, Y., & Kuh, D. (2002). A life course approach to chronic disease epidemiology: Conceptual models, empirical challenges and interdisciplinary perspectives. *International Journal of Epidemiology, 31*(2), 285-293. https://doi.org/10.1093/ije/31.2.285

Bishedsari, F., Mallekzadeh, R., Maldavinia, M., Vacca, M., & Oariani-Costantini, R. (2014), Epidemiological transition of colorectal cancer in developing countries:

Environmental factors. Molecular pathway, and opportunity for prevention. *World Journal of Gastroenterology, 20*(20), 6055-6072. https://doi.org/10.3748/wjg.v20.i20.6055

Blane, D., Netuveli, G., & Stone, J. (2007). The development of life course epidemiology. *Revue d'Épidémiologie et de Santé Publique 55*(1), 31-38. https://doi.org/10.1016/j.respe.2006.12.004.

Bodo, T., Gimah, B. G., & Ajuru, I. (2019). Oil crisis in the Niger Delta Region of Nigeria: Genesis and Extent. *European Scientific Journal, 15 (36), 1857-7881*. http://doi.org/10.19044/esj.2019.v15n36p141

Boo, S., & Froelicher, S. (2013). Secondary analysis of national datasets. *Japan Journal of Nursing Science, 10,* 130-135. https://doi.org/10.1111/j.1742-7924.2012.00213.x.

Boscoe, F. P., Johnson, C. J., Sherman, R. L., Stimchcomb, M. S., Lim, G & Henry, K. A. (2014). The relationship betweenarea poverty rate and site-specific cancer incidence in the United States. *Cancer*. https://doi.org/10.1002/cmcr.28632.

Bray, F., Ferlay, J., Soerjomataran, J., Siegel, R. L., Torre, L. A., & Jenal, A. (2018). Global cancer statistics: GLOBOCAN estimates of incidence and mortality worldwide for 36cancers in 195 countries. *CA: A Cancer Journal for Clinicians, 68*(6), 394-424. https://doi.org/10.3322/caac.21492

Burton-Jeangros. C., Cullati, S., Fribourg, U., & Sacker, A., & Bane D. (2015). A life course perspective on health trajectories and transitions. *Springer Open, 4*, 19-197. https://doi.org/DOI 10.1007/978-3-319-20484-0

Calkins, K., & Devaskar, S. U. (2011). Foetal orign of adult diseases. Curr ProbllPediatr Adolesc Health Care, 41 (6), 158-171. https://doi.org/10.1016/j.cppeds.2011.01.001

Canchola, A. J., Marco, S. S., Yang, J., Hertz, A., Park, S., Albright, C., …& Cheng, I. (2017). Association between the neighborhood obesogenic environment and colorectal cancer risk in the multiethnic cohort. *Cancer Epidemiology, 50*, 99-106. https://doi.org/10.1016/jcanep.2017.08.009.

Cavalin, C. (2019). Encyclopedia of the environment. https://www.encyclopedie-environnement.org/en/society/exposome-social-sciences-promise-of-meeting/

Centers for Disease Control and Prevention. (2018a). Confidentiality. Retrieved 30th July 2020 from https://www.cdc.gov/nchs/nhanes/participant/participant-confidentiality.htm

Chao, A., Thun, M. J., Jacobs, E. J., S. Henley, S. J.,Rodriguez, C., & Calle, E.E.(2000). Cigarette Smoking and Colorectal Cancer Mortality in the Cancer Prevention Study II. *Journal of the National Cancer Institute*, 92 (23),1888–1896. https://doi.org/10.1093/jnci/92.23.1888

Chen, K., Yu, W., Ma, X., Yao, K., & Jiang, Q. (2005). The association between drinking water source and colorectal cancer incidence in Jiashan County of China: A prospective cohort study. *European Journal of Public Health, 15*(6), 652-656. https://doi.org/10.1093/eurpub/cki027

Cheng, H. G., & Phillips, M. R. (2014). Secondary analysis of existing data: opportunities and implementation. *Shanghai Archives of Psychiatry*, *26*(6), 371–

375. http://doi.org/10.11919/j.issn.1002-0829.214171

Chikwem, F.C., & Duru, J. C. (2018). The Resurgence of the Niger Delta militants and the survival of the Nigerian state. The Commonwealth *Journal of International Affairs,107(*1),45-55. https://doi.org/10.1080/00358533.2018.1424074

Chinedu, E., & Chukwuemeka, C. K. (2018). Oil spillage heavy metals toxicity risk in the Niger Delta. *Journal of Health and Pollution, 8*(19), 1-8. https://doi.org/10.5696/2156-9614-8.19.180905.

Chow, W. H., Malker, H. S., Hsing, A. W., McLaughlin, J. K., Weiner, J. A., Stone, B. J., Ericsson, J. L., & Blot, W. J. (1994). Occupational risks for colon cancer in Sweden. *Journal of Occupational Medicine, 36*(6), 647-651. Article 8071728.

Chu, H., Xin, J., Yuan, Q, Wu, Y., Du, M., Zheng, R., Liu, H., Wu, S., Zhang, Z., & Wang, M. (2020). A prospective study of the association among fine particulate matter, genetic variants, and the risk of colorectal cancer. *Environmental International, 147* (106309). https://doi.org/10.1016/j.envint.2020.106309

Clarke, S., & Cossette, S. (2000). Secondary analysis: Theoretical, methodological, and practical considerations. *Canadian Journal of Nursing Research, 32*, 109-129. https://doi.org/pubmed.ncbi.nlm.nih.gov/11928128

Clarke, M. A., & Joshu, C. E. (2017). Early life exposure and adult cancer risks. Epidemiologic Review, 36(1), 11-27. https://doi.org/10.1093/epirev/mxx004

Colorectal Cancer Collaborators. (2019). The global, regional, and national burden of colorectal cancer and its attributable risk factors in 195 countries and territories, 1990-2017: A systematic analysis for the global burden of disease study 2017.

The Lancet, 4(12). https://doi.org/10.1016/S2468-1253(19)30345-0

Colussi, D., Fabbri, M., Zagari, R. M., Montale, A., Bazzoli, F., & Ricciardiello, L. (2018). Lifestyle factors and risk for colorectal polyps and cancer at index colonoscopy in a FIT-positive screening population. *United European gastroenterology journal*, *6*(6), 935–942. https://doi.org/10.1177/2050640618764711

Conrad, P., & Barker, K. K. (2010). The social construction of illness: Key insights and policy implications. *Journal of Health and Social Behavior, 51*(S), S57-S79. https://doi.org/10.1177/0022146510383495.

Creswell, J. W., Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods* (5th ed.). Sage Publications. https://onlinelibrary.wiley.com/doi/full/10.1002/nha3.20258

D'Andrea, M. A., & Reddy, G. K. (2018). The development of long-term adverse health effects in oil spill clean-up workers of the Deepwater horizon offshore drilling rig disaster. *Frontiers Public Health, 6,* 117. https://doi.org/10.3389/fpubh.2018.00117.

Daniel, W. W., & Cross, C. L. (2013). *Biostatistics: A Foundation for Analysis in the Health Sciences*. Hoboken, NJ: Wiley. https://www.amazon.com/Biostatistics-Foundation-Analysis-Health-Sciences/dp/1118302796

Dashti, S. G., Buchanan, D. D., Jayasekara, H., Ouakrim. D. A., Clendenning, M., Rsoty, C., Winship, I. M., Macrea, F. A., Giles, G. G., Parry, S., Casey, F. A., Haile. R. W., Gallinger, S., Marchand, L. L., Thibodeau, S. N., Lindor, N. M., Newcomb,

P. A., Potter, J. D., Baron, J. A., Hopper, J. L., …& Win A. K. (2017). Alcohol Consumption and the Risk of Colorectal Cancer for Mismatch Repair Gene Mutation Carriers. *Cancer Epidemiol Biomarkers Prev*ention, 26(3), 366-375. https://doi.org/doi:10.1158/1055-9965.EPI-16-0496

Davis-Kean, P. E., Jager, J., & Maslowsky, J. (2015). Answering developmental question using secondary data. *Child Development Perspective, 9(*4), 256-261. https://doi.org/10.1111/cdep.12151

De la Sablonniere, R. (2017). Towards a psychology of social change: A typology of social change front. *Psychology 8(*397). https://doi.org/10.3389/fpsyg.2017.00397

Dinis-Oliveira, R. (2016). Oxidative and Non-Oxidative Metabolomics of Ethanol. *Current Drug Metabolism, 17(*4):327-335. https://doi.org/10.2174/1389200217666160125113806.

Diez-Roux, A. V. (2000). Multilevel analysis of public health research. *Annual Review of Public Health*, *21*, 171-192. https://doi.org/10.1146/annurev.publhealth.21.1.171.PMID:10884951

Doolan, D. & Froelicher, E. (2009). Using an existing data set to answer new research questions: A methodological review. *Research and Theory for Nursing Practice*, *23*, 203–215. http://dx.doi.org/10.1891/1541-6577.23.3.203

Doubeni, C. A., Laiyemo, O., Major, J. M., Schootman, M., Park, Y., Graubard, B. L.,… & Sinha, R. (2012). Socioeconomic status and the risk of colorectal cancer: An analysis of over one-half million adults in the NIH-AARP diet and health study. *Cancer, 118*(14), 3636-3644. https://doi.org/10.1002/cncr.26677.

Dumas, S., Parent, M. E., Siemiatycki, J., Brisson, J. (2000). Rectal cancer and occupational risk factors: A hypothesis-generating, exposure-based case–control study. *International Journal of Cancer, 87*: 874-879. https://doi.org/10.1002/1097-0215(20000915)87:6<874: AID-IJC18>3.0.CO;2-L.

El-Zaemey, S., Anand, T, N., Heyworth, J. S., Boccupational oyle, T., van Tongeren, M., & Lin, F. (2018). Case-control study to assess the association between colorectal cancer and selected occupational agents using INTEROCC job exposure matrix. *Occupational and Environmental Medicine, 75(*4), 290. http://dx.doi.org/10.1136/oemed-2017-104795

Ethan, C. J., Mokoena, K. K., Yu, Y., Shale, K., Fan, Y., Rong, J., Liu, F. (2020). Association between $PM_{2.5}$ and mortality of stomach and colorectal cancer in Xi'an: a time-series study. *Environmental Science Pollution Research International 27*(18):22353-22363. https://doi.org/10.1007/s11356-020-08628-0.

Elum, Z. A., Mopipi, K., & Henri-Ukoha, A. (2016). Oil exploration and its socioeconomic effects on the Niger-Delta region of Nigeria. *Environmental Science and Pollution Research, 23(*13), 12880-12889. https://doi.org/10.1007/s11356-016-6864-1.

El-Zaemey, S., Anand, T, N., Heyworth, J. S., Boccupational oyle, T., van Tongeren, M., & Lin, F. (2018). Case-control study to assess the association between colorectal cancer and selected occupational agents using INTEROCC job exposure matrix. *Occupational and Environmental Medicine, 75(*4), 290. http://dx.doi.org/10.1136/oemed-2017-104795.

Fairley, T. L, Cardinez, C. J., Martin, J., Alley. L., Friedman, C., Edwards, B., & Jamiso, P. (2006). Colorectal cancer in U.S. adults younger than 50 years of age, 1998–2001. *Cancer,*107(5, Suppl):1153–61. https://doi.org/10.1002/cncr.22012.

Fann, N., Baker, K. R., Chan, E., Eyth, A., Macpherson, A., Miller, E., & Snyder, J. (2018). Assessing Human Health $PM_{2.5}$ and Ozone Impacts from U.S. Oil and Natural Gas Sector Emissions in 2025. *Environmental science & technology*, *52*(15), 8095–8103. https://doi.org/10.1021/acs.est.8b02050.

Fano, V., Cernigliaro, A., Scondotto, S., Addario, S. P., Caruso, A., Mira, A., Forastiere, F., & Perucci, C. A. (2006). Mortality (1995–2000) and hospitalizations (2001–2003) in the industrial area of Gela (in Italian). *Preventive Epidemiology, 30 (*1), 27-32. https://doi.org/pubmed.ncbi.nlm.nih.gov/16826696

Faul, F., Erdfelder, E., Lang, A.G., & Axiel, B (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioural, and biomedical sciences. *Behavior Research Methods* 39, 175–191. https://doi.org/10.3758/BF03193146.

Federal Ministry of Health. (2014). *National Malaria Elimination Strategic Plan (*2014-2020). Federal Ministry of Health of Nigeria. https://health.gov.ng/doc/NMEP-Strategic-Plan.pdf

Ferlay, J., Soerjomataram, J. I.,Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Bray, F. (2015). Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *International Journal of Cancer. 136*(5), E359-

E386. https://doi.org/10.1002/ijc.29210.

Freedman, V. A., Grafova, I. B., & Rogowski, J. (2011). Neighborhoods and chronic disease onset in later life. *American journal of public health*, *101*(1), 79–86. https://doi.org/10.2105/AJPH.2009.178640

Gabriel, E., Attwood, K., Al-Sukhni, E., Erwin, D., Boland, P., & Nurkin, S. (2018). Age-related rates of colorectal cancer and the factors associated with overall survival. *Journal of gastrointestinal oncology*, *9*(1), 96–110. https://doi.org/10.21037/jgo.2017.11.13

Godfrey, K. M., & Barker, D. J. P. (2000). Fetal nutrition and adult disease. *American Journal of Clinical Nutrition, 71(*S), 1344S-1352S. https://doi.org/10.1139/Y09-006

Goldberg, M. S., Parent, M. E., Siemiatycki, J., Désy, M., Nadon, L., Richardson, L., Lakhani, R., Latreille, B., Valois, M. F. (2001). A case–control study of the relationship between the risk of colon cancer in men and exposures to occupational agents. *American Journal of Industrial Medicine, 39*(6), 531-546. https://doi.0rg/10.1002/ajim.1052.

Grandos-Romero, J. J., Valderrama-Treino, A. I., Contreras-Flores, E. H., & Barrera-Mera, B., Emriquez, M. H., Uriarte-Ruiz, K., Ceballos-Villalva, C. C., Estrada-Mata, A. C., Rodriquez, C. A., & Aranz-Peria, G. (2017). Colorectal cancer: A review. *International Journal of Research in Medical Sciences, 5(*11), 4667-4676. http://dx.doi.org/10.18203/2320-6012.ijrms20174914

Guo, C., Chan, T. C., Lin, C., Bo, Y., Chang, L. Y., Lau, A. K. H., Wong, M. C. S., &

Qian, L. X. (2020). Long-term exposure to ambient fine particles and gastrointestinal cancer mortality in Taiwan: A cohort study. *Environment International, 138* (105640). https://doi.org/10.1016/j.envint.2020.105640.

Hajat, A., Hsia, C., & O'Neill, M. S. (2015). Socioeconomic Disparities and Air Pollution xposure: A Global Review. *Current environmental health reports*, *2*(4), 440–450. https://doi.org/10.1007/s40572-015-0069-5

Hannan, L. M., Jacobs, E. J., & Thun, M. J. (2009). The association between cigarette smoking and risk of colorectal cancer in a large prospective cohort from the United States. Cancer Epidemiology, Biomarkers and Prevention, 16(12), 3363-3367. https://doi.org/10.1158/1055-9965.EPI-09-0661

Hatch, M., & Thomas, D. (1993). Measurement issues in environmental epidemiology. *Environmental Health Perspective, 101*(Suppl. 4), 49–57. https://doi.org/10.1289/ehp.93101s449

Hurley, S., Goldberg, D., Nelson, D. O., Lu, Y., Henderson, K., Bernstein, L., & Reynolds, P. (2013). Risk of colorectal cancer associated with active smoking among female teachers. *Cancer causes & control: CCC*, *24*(7), 1291–1304. https://doi.org/10.1007/s10552-013-0207-z

Hurtado, J. L., Bacigalupe, J. A., Calvo, M., Esnaola, S., Mendizabal, N., Portillo, I.,…& Arana-Arri E. (2015). Social inequality in a population based colorectal in the Basque country. *BMC Public Health, 15*(1021), 2-14. https://doi.org/10.1186/s12889-015-2370-5

Ikoba, N. A., Jolayemi, E. T., & Sanni, O. O. M. (2018). Nigeria's recent population

censuses: A Benford-theoretic evaluation. *African Population Studies, 32*(1), 3974-3981. https://doi.org/10.11564/32-1-1166

IQAir (2021). Air quality index (AQI) and PM2.5 air pollution in Nigeria. **https://www.iqair.com/us/nigeria**

Irabor, O. O., Atuwapo, O. O., & Ayandipo, O. O. (2014). The present study of the management of colon and rectal cancers in Nigeria. *Journal of Cancer Research, 2014*. https://doi.org/10.1155/2014/267190.

Irabor, D. A. (2018). Emergence of colorectal cancer in West Africa: Accepting the Inevitable. *Nigerian Medical Journal, 58*(3), 87-91. https://doi.org/10.4103/0300-1652.234076

Inoni, O., Omotor, O., & Nkem, A. F., (2006). The effect of oil spillage on crop yield and farm income in Delta State, Nigeria. *Journal of Central European Agriculture, 7*(1), 41-48. https://doi.org/10.5513/jcea.v7i1.353.

Jeddy-Agba, E., Curado, M. P., Ogunbiyi, O., Oga, E., Fabowale, T., Igbiniba, F., &…Adebamowo, C. A. (2012). Cancer incidence in Nigeria: A report from population-based cancer registries. *Cancer Epidomiology, 36*(5), 271-278. https://doi.org/10.1016/j.canep.2012.04.007.

Jenwitheesuk, K., Peansukwech, U., & Jenwitheesuk, K. (2020). Accumulated ambient air pollution and colon cancer incidence in Thailand. *Scientific Report, 10* (17765). https://doi.org/10.1038/s41598-020-74669-7

Jin, M., Cal, S., Guo, J., Zhu, Y., Li, M., Yu, Y., Zhang, S., & Chen, K. (2013). Alcohol drinking and all cancer mortality: A meta-analysis. Annals of Oncology, 23(3).

https://doi.org/10.1093/annonc/mdq653

Johnson, N. P. (2014). Secondary Data Analysis: A Method of which the time has come. Qualitative *and Quantitative Methods in Libraries (QQML) 3*, 619 –626.

Johnston, J. E., Lim, E., & Roh, H. (2019). Impact of upstream oil extraction and environmental pollution: *A review of the evidence. Science of the Total Environment, 657,* 187-199. https://doi.org/10.1016/j.scitotenv.2018.11.483.

Jolliffe. I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical and Transactional Research, 374* (2065), 1-16. http://doi.org/10.1098/rsta.2015.0202

Kachuri, L., Villeneuve, P. J., Parent, M., Johnson, K. C., & Harris, S. A. (2016). Workplace exposure to diesel and gasoline engine exhaust and the risk of colorectal cancer in Canadian men. *Environmental Health, 15*(4), 2-12. https://doi.org/10.1186/s12940-016-0088-1

Kashino, I., Mizoue, t., tanaka, K., Tsuji, I., Tamakoshi, A., Matsuo, K., Wakai, K., Nagata, C., Inoue, M., Tsugane, S., Sasasuki, S., Iwakashi,M., Otani, T., Sawada, N., Tsuji, Y., Tsubonio, Y., Nishino, Y., Ito, H., Wakai, K. (2015). Vegetable consumption and colorectal cancer risk: An evaluation based on a systematic review and meta-analysis among the Japanese population, *Japanese Journal of Clinical Oncology*, 45 (10),973–979. https://doi.org/10.1093/jjco/hyv111

Khallaf, M. (2011). The Impact of Air Pollution on Health, Economy, Environment and Agricultural Sources. Intect Open. https://doi.org/10.5772/1000.

Klein, N. (2017). *No is not enough: Resisting Trump's shock politics and winning the world we need.* Haymarkel Books. https://doi.org/10.1186/s12889-015-2370-5

Kikuchi, N., Nishiyama, T., Sawada1, T., Wang, C., Lin, Y., Watanabe, Y.,…& Kikuchi, S. (2017). Perceived Stress and colorectal cancer incidence. The japan collaborative cohort study. *Scientific Report, 7*, 40363. https://doi.org/10.1038/srep40363

Kim, D., Masyn, K. E., Kawachi, I., Laden, F., & Colditz, G. A. (2010). Neighborhood socioeconomic status and behavioral pathways to risks of colon and rectal cancer in women. *Cancer*, *116*(17), 4187–4196. https://doi.org/10.1002/cncr.25195)----- DONE

Krieger, N. (1994). Epidemiology and the web of causation: Has anyone seen the spider? *Social Science & Medicine, 39*, 887–903. https://doi.org/10.1016/0277-9536(94)90202-X

Krieger, N. (2011). *Epidemiology and the people's health: Theory and context.* Oxford University Press. Article number 142957689. https://doi.org/10.5860/choice.49-0897

Krieger, N. (2013). History, biology, and health inequalities: Emergent embodied phenotypes and the illustrative case of the breast cancer estrogen reception. *American Journal of Public Health, 103(*1), 22-27. https://doi.org/10.2105/AJPH.2012.300967.

Krieger, N., Waterman, P. D., Spasojevic, J, Li. W, Maduro, G., & Van Wye, G. (2016). Public health monitoring of privilege and deprivation with the index of

concentration at the extremes. *American Journal of Public Health, 106*(2), 256–263. https://doi.org/10.2105/AJPH.2015.302955.

Krieger, N., Jahn, J. L., Waterman, P. D., & Chen J. T. (2018). Breast cancer estrogen receptor status according to biological generation. US black and white women born 1915-1979. *American Journal of Epidemiology, 187*(5), 960–970. https://doi.org/10.1093/aje/kwx312.

Kruk J., Aboul-Enein, B. H., Bernstein, J., Gronostai, M. (2019). Psychological Stress and Cellular Aging in Cancer: A meta-Analysis. *Oxidative Medicine and Cellular Longevity, 13*, 2-21. https://doi.org/10.1155/2019/1270397.

Kuh D., & Shlomo, Y. B. (2004). *Introduction to a life course approach to chronic disease epidemiology.* Oxford University Press.Blane, D., Netuveli, G., Stone, J. (2007). The development of life course epidemiology. *Revue d'Épidémiologie et de Santé Publique, 55*(1). 31-38. https://doi.org/10.1016/j.respe.2006.12.004

Kunzmann, A. T., Coleman, H. G., Huang, W. Y., Cantwell, M. M., Kitahara, C. M., & Berndt, S. I. (2016). Fruit and vegetable intakes and risk of colorectal cancer and incident and recurrent adenomas in the PLCO cancer screening trial. *International journal of cancer*, *138*(8), 1851–1861. https://doi.org/10.1002/ijc.29922

Lancaster, G. A, Green, M. & Lane, S. (2006). Reducing bias in ecological Studies: An evaluation of different dethodologies. *Journal of the Royal Statistical Society, 169*(4), 681-700. http://www.jstor.org/stable/3877395

Lee, C., Famoye, F., Shelden, B. & Brown, A. (n.d.a). Data manipulation, SPSS On-Line Training Workshop. Retrieved from

http://calcnet.mth.cmich.edu/org/spss/data_manipulation.htm

Lopez-Abente, G., Garcia-Perez, J., Fernandez-Navarro, P., Boldo, E., & Ramis, R. (2012). Colorectal cancer mortality and industrial pollution in Spain. BMC Public Health, 12 (589), 2-12. doi: 10.1186/1471-2458-12-589. https://doi.org/10.1186/1471-2458-12-589.

Li, K., Lin, G. Z., Li, Y., Dong, H., Xu, H., Song, S., Liang, Y., & Liu, H. (2017). Spatio-temporal analysis of the incidence of colorectal cancer in Guangzhou, 2010–2014. *Chinese Journal of Cancer 36 (*60), 1-8. https://doi.org/10.1186/s40880-017-0231-6

Lin, J., Zhang, S. M., Cook, N. R., Rexrode, K. M., Liu, S., Manson, J. E., Lee, I. M., & Buring, J. E. (2005). Dietary Intakes of Fruit, Vegetables, and Fiber, and Risk of Colorectal Cancer in a Prospective Cohort of Women (United States). *Cancer Causes & Control, 16(3), 225-233.* https://www.jstor.org/stable/20069461.

Marley, A. R., & Nan, H. (2016). Epidemiology of colorectal cancer. *International Journal of Molecular Epidemiology and Genetics, 7(*3), 105-114. https://pubmed.ncbi.nlm.nih.gov/27766137/

Marmot, M. G. (2014). *The status syndrome: How social standing affects our health and longevity.* Holt Publishing. https://www.barnesandnoble.com/w/status-syndrome-michael-marmot/1100625927

Moses, D., & Olaniyi, A. T. (2017). Resurgence of Militancy in the Niger Delta Region of Nigeria. *Journal of political Science Affairs, 5*: 298. https://doi.org/10.4172/2332- 0761.1000298.

Nriagu, J., Udofia, H. A., Ekong, I., & Eduk, G. (2016), Health risks associated with oil pollution in the Niger Delta, Nigeria. *International journal of Environmental Research and Public Health, 13(*346), 1-23. https://doi.org/10.3390/ijerph13030346.

Mansori, K., Solaymani-Dodaran, M., Mosavi-Jarrahi, A., Motlagh, A. G., Salehi, M., Delavari, A., & Asadi-Lari, M. (2018). Spatial Inequalities in the Incidence of Colorectal Cancer and Associated Factors in the Neighborhoods of Tehran, Iran: Bayesian Spatial Models. *Journal of preventive medicine and public health, 51*(1), 33–40. https://doi.org/10.3961/jpmph.17.167.

Marie-Élise, P., Marie-Claude, R., Paolo, B., Aaron, C., & Jack, S. (2007). Exposure to Diesel and Gasoline Engine Emissions and the Risk of Lung Cancer, *American Journal of Epidemiology*, *165* (1), 53–62, https://doi.org/10.1093/aje/kwj343

Marmot, M. G. (2014). *The status syndrome: How social standing affects our health and longevity.* Holt Publishing. https://www.barnesandnoble.com/w/status-syndrome-michael-marmot/1100625927

Miles, D.A. (2017) Workshop: Confessions of a Dissertation Chair Part 1: The Six Mistakes Doctoral Students Make with the Dissertation. Presented at the 5th Annual 2017 Black Doctoral Network Conference in Atlanta, GA on October 26-29, 2017. https://www.researchgate.net/publication/318043744_CONFERENCE_WORKSHOP_Confessions_of_a_Dissertation_Chair_The_Six_Mistakes_Doctoral_Students_Make_With_the_Dissertation

Mills, P., & Yang, R. (1997). Cancer incidence in the Hmong of central California, United States, 1987-1994. *Cancer Causes Control, 8*, 705-712. https://doi.org/10.1023/a:1018423219749.

Moore, K. J., Sussman, D. A., & Koru-Sengul, T. (2018). Age-Specific Risk Factors for Advanced Stage Colorectal Cancer, 1981-2013. *Preventing chronic disease*, *15*, E106. https://doi.org/10.5888/pcd15.170274.

Moses, D., & Olaniyi, A. T. (2017). Resurgence of Militancy in the Niger Delta Region of Nigeria. *Journal of political Science Affairs, 5*: 298. https://doi.org/10.4172/2332- 0761.1000298.

National Population Commission (NPC) [Nigeria] and ICF. 2019. *Nigeria Demographic and Health Survey 2018*. Abuja, Nigeria, and Rockville, Maryland, USA: NPC and ICF**.** https://nigeriahealthwatch.com/wp-content/uploads/bsk-pdf-manager/2019/12/NDHS-2018.pdf

Nnaemeka, A. N. (2020). Environmental pollution and associated health hazard to host communities (Case study: Niger-Delta Region of Nigeria). *Central Asian Journal of Environmental Science and Technology Innovation, 1*(2020), 30-42. https://doi.org/ 10.22.22034/CAJESTI.2020.01.04.

Nnagu, J., Udofia, H. A., Ekong, I., & Eduk, G. (2016), Health risks associated with oil pollution in the Niger Delta, Nigeria. *International journal of Environmental Research and Public Health, 13(*346), 1-23. https://doi.org/10.3390/ijerph13030346.

Nnodim, I. (2021). Nigeria lost N898.93bn to pipeline vandalization, theft in nine months. Punch Newspaper. https://punchng.com/nigeria-lost-n898-93bn-to-pipeline-vandalism-theft-in-nine-months/

Nriagu, J., Udofia, H. A., Ekong, I., & Eduk, G. (2016), Health risks associated with oil pollution in the Niger Delta, Nigeria. *International journal of Environmental Research and Public Health, 13(*346), 1-23. https://doi.org/10.3390/ijerph13030346.

Obiorah, C. C., & Nwafor, C. C. (2019). Cancer mortality in the Niger Delta Region of Nigeria: A case study of the University of Port Harcourt Teaching Hospital. *Nigerian Medical Journal, 60(*5), 268-272. https://doi.org/10.4103/nmj.NMJ_51_19.

Oddone, E., Modonesi, C., Gatta, G. (2014). Occupational exposures and colorectal cancers: A quantitative overview of epidemiological evidence. *World Journal of. Gastroenterology, 20*, 12431e12444. https://doi.org/10.3748/wjg.v20.i35.12431

Oghenetaga, O. B., Ana, R. E. E., Okunola, M. A., & Ojengbade, O. A. (2020). Oil spills, gas flaring, and adverse pregnancy outcome: A systematic Review. Open Journal of Obstetrics and Gynaecology, 10, 187-199. https://doi.org/10.1016/j.scitotenv.2018.11.483.

Ordinioha, B., & Brisibe, S. (2013). The human implication of oil spillage in Niger delta Nigeria: An interpretation of published studies. *Journal of the Nigeria Medical Association, 54(*1), 10-16. https://doi.org/10.4103/0300-1652.108887.

Osuagwu, E. S., & Olaifa, E. (2018). Effects of oil spills on fish production in the Niger

Delta. *PloS One, 13*(10), 1-14. https://doi.org/10.1371/journal.pone.0205114.

Parajuli, R., Bjerkaas, E., Tverdal, A.., Marchand, L. L., Weiderpass, E., & Gram, I. T. (2014). Smoking increases rectal cancer risk to the same extent in women as in men: results from a Norwegian cohort study. *BMC Cancer* 14, 321. https://doi.org/10.1186/1471-2407-14-321

Park, Y., Subar, A. F., Kipnis, V., Thompson, F. E., Mouw, T., Hollenbeck, A., Leitzmann, M. F., & Schatzkin, A. (2007). Fruit and Vegetable Intakes and Risk of Colorectal Cancer in the NIH–AARP Diet and Health Study, *American Journal of Epidemiology*, *166*(2), 170–180. https://doi.org/10.1093/aje/kwm067

Pascal, L., Pascal, M., Stempfelet, M., Goria, S., & Declerco, C. (2013). Ecological study on hospitalization for cancer, cardiovascular, and respiratory diseases in the industrial area of Etanf-de-Berre in the South of France. *Journal of Environmental and Public Health.* https://doi.org/10.1155/2013/328737

Peters, H., Hoek, G., & Ketsouyanni, K. (2012). Understanding the link between environmental exposure and health: Does the Exposome promise too much? *Journal of Epidemiology and Community Health, 66*, 103-105. https://doi.org/10.1136/jech-2011-200643

Polit, D. (2010). *Statistics and data analysis for nursing research* (2nd edn). Upper Saddle River, NJ: Pearson Prentice Hall Raman.

Ray Ofor, E. G., Gbaanador, B. M., Madubuike, C. O., Obiorah, C. C., & Jebbin, N. J. (2020). Colorectal cancer in Niger Celta, Nigeria: Multi-centre lower

gastrointestinal endoscopy study. *Journal of Medical Research and Case Report, 2*(1). 1-5. http://doi.org/10.5281/zenodo.3668422

Rocco, T. S. (2009). Literature review, conceptual framework, theoretical framework, terms, functions, and distinctions. *Human Resource Development, 8(*1), 129-130. https://doi.org/10.1177/1534484309332617

Rogala, D., Marchwinska-Wyrwal, E., Spychala, A., & Hajok, I. (2019). Incidence of colorectal cancer in urban population exposed to Cadmium. *Polish Journal of Environmental Studies, 28(*5), 3395-3400. https://doi.org/10.15244/pjoes/9270

Rosenberg, M. S. (2000). The bearing correlogram: A new method of analyzing directional spatial autocorrelation. *Geographical Analysis, 32*, 267–278. https://doi.org/ 10.1111/j.1538-4632. 2000.tb00428. x.

Rossi, M., Jahanzaib, A., Usman, A., Keshavarzian, A., & Bishehsari, F. (2018). Colorectal Cancer and Alcohol Consumption-Populations to Molecules. *Cancers*, *10*(2), 38. https://doi.org/10.3390/cancers10020038

Rudestam, K. E., & Newton, R. R. (2015). *Surviving your dissertation: A comprehensive guide to content and process* (4th ed.). Thousand Oaks, CA: Sage. https://eric.ed.gov/?id=ED500420

Sato, Y., Tsubono, Y., Nokava, N., & Ogawa, K. (2007). Fruit and vegetable consumption and risk of colorectal cancer in Japan: The Miyagi Cohort Study. Public Health Nutrition, 8(3), 309-314. https://doi.org/10.1079/PHN2004681

Schwartz, S. (1994). The fallacy of ecological fallacy: The potential misuse of the concept and the consequences. *American Journal of Public Health, 84*(5), 819-

824. https://doi.org/10.2105/AJPH.84.5.819.

Sedgwic. P. (2014). Ecological studies: Advantages and disadvantages. *British Medical Journal, 348,* 2979. https://doi.org/10.1136/bmj.g2979

Setia, M. S. (2016). Methodology Series Module 3: Cross-sectional Studies. *Indian Journal of Dermatology*, *61*(3), 261–264. http://doi.org/10.4103/0019-5154.182410.

Shah, S. A., Neoh, H., Shariman, S., Rahim, S., Azhar, Z. I., Hassan, M. R., Safian, N., & Jamal, R. (2014). Spatial analysis of colorectal cancer cases in Kuala Lumpur. *International Journal of Health Geographics, 6(*1), 1-10. http://doi.org/10.7314/apjcp.2014.15.3.1149

Shiao, S., Grayson, J., Yu, C. H., Wasek, B., & Bottiglieri, T. (2018). Gene Environment Interactions and Predictors of Colorectal Cancer in Family-Based, Multi-Ethnic Groups. *Journal of personalized medicine*, *8*(1), 10. https://doi.org/10.3390/jpm8010010

Siegel, R. L., Fedewa, S. A., Anderson, W. F., Miller, K. D., Ma, J., Rosenberg, P. S., & Jemal, A. (2017). Colorectal cancer incidence patterns in the United States, 1974–2013. *Journal of National Cancer Institute,109*(8), djw322. https://doi.org/10.1093/jnci/djw322.

Siopen, N., Kontos, E. Z., Ryff, C. D., & Ayanian, Y. Z. (2013). Psychosocial stress and cigarette smoking, persistence, cessation, and relapse over 9-10 years: A prospective study of diddle-aged adults in the United States. *Cancer Causes and Control, 24*(10), 1849-1863. https://doi.org/10.1007/s10552-013-0262-5.

Slattery, M. L. (2004). Physical Activity and Colorectal Cancer. *Sports Med* 34, 239–252. https://doi.org/10.2165/00007256-200434040-00004.

Society of General Internal Medicine. (2012). *User's guide to working with secondary data*. http://www.sgim.org/communities/research/dataset-compendium/users-guide-to-working-with-secondary-data

Song, N., Lee, J., Cho, S., Kim, J., Oh, J. H., & Shin, A. (2019). Evaluation of gene-environment interactions for colorectal cancer susceptibility loci using case-only and case-control designs. *BMC Cancer 19*, 1231. https://doi.org/10.1186/s12885-019-6456-9

Statement of Ethical Practice for the British Sociological Association. (2004). The British Sociological Association, Durham. http://www.york.ac.uk/media/abouttheuniversity/governanceandmanagement/governance/ethicscommittee/hssec/documents/BSA%20statement%20of%20ethical%20practice.pdf

Stevenson, M., & McClure, R. (2005). Study design for injury prevention. *BMJ, 1*(1), 14. http://dx.doi.org/10.1136/ip.2003.003384

Stingone, J. A., Buck Louis, G. M., Nakayama, S. F., Vermeulen, C.H., Kwok, R. K., Cui, Y., Balshaw, D. M., &Teitelbaum, S. L. (2017). Toward Greater Implementation of the Exposome Research Paradigm within Environmental Epidemiology. *Annual Review of Public Health*, *38,* 315-327. https://doi.org/10.1146/annurev-publhealth-082516-012750

Su, S. Y., Liaw, Y. P., Jhuang, J. R., Hsu, J., Chiang, C., Yang, Y., & Lee, W.

(2019). Associations between ambient air pollution and cancer incidence in Taiwan: an ecological study of geographical variations. *BMC Public Health* 19, 1496. https://doi.org/10.1186/s12889-019-7849-z

Tavolacci, M.P., Ladner, J., Grigioni, S., Richard, L., Ville, H., & Dechelotte, P. (2013). Prevalence and association of perceived stress, substance use and behavioral addictions: a cross-sectional study among university students in France, 2009–2011. *BMC Public Health* 13, 724 (2013). https://doi.org/10.1186/1471-2458-13-724

Thune, I., & Furberg, A. S. (2001). Physical activity and cancer risk: Dose-response and cancer, all sites and site-specific. *Medicine in Science and Sports and Exercise*. 33(6 Suppl), S530-S610. https://doi.org/10.1097/00005768-200106001-00025

Tripathy, J. P. (2013). Secondary data analysis: Ethical issues and challenges. *Iranian Journal of Public Health, 42(*2), 1478-1479. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4441947/

Turner, M. C., Krewski, A., Oliver, W. R., Pope, C. A., Burnett, R. T., Jerrrell, M.,…Gapstur, S. M. (2017). Ambient air pollution and cancer mortality in the cancer promotion study. Environmental Health Perspectives, 125(8). https://doi.org/10.1289/EHP1249

Udotong, J. I. R., Udoudo, U. P., & Udotong, L. R. (2017). Effects of oil and gas exploration and production activities on the production and management of seafood in Akwa Ibom state, Nigeria. *Journal of Environmental Chemistry and Ecotoxicology, 9*(3), 20-42. https://doi.org/10.5897/JECE2017.0405

Valle, L. (2014). Genetic predisposition to colorectal cancer: where we stand and future perspectives. *World journal of gastroenterology*, *20*(29), 9828–9849. https://doi.org/10.3748/wjg.v20.i29.9828.

Vargas, G. G., Au, W. W., & Izotti, A. (2020). Public Health issues from crude-oil production in the Ecuadorian Amazon territory. *Science of the Total Environment, 719*, 1-9. https://doi.org/10.1016/j.scitotenv.2019.134647

Vega, P., Valentin, F., & Cubiella, J. (2015). Colorectal cancers diagnosis: Pitfalls and opportunities. *World Journal of Gastroenterology Oncology, 7(*12), 422-433. https://doi. org/10.4251/wjgo.v7.i12.422.

Ventriglio, A., Bellomo, A., & Bhugra, D. (2015). Webs of causation and its implication for epidemiological research. *International Journal of Social Psychiatry, 62*(1), 3-4. . https://doi.org/10.1177/0020764015587629

Viatte, C., Clerbaux, C., Maes, C., Daniel, P., Garello, R., Safieddine, S/., & Ardhuin, F. (2020). Air Pollution and Sea Pollution Seen from Space. *Survey Geophysics 41*, 1583–1609 (2020). https://doi.org/10.1007/s10712-020-09599-0

Vogtmann, E., Xiang, Y. B., Li, H. L, Levitan, E. B., Yang, G., Waterbor, W., Gao, J., Cai, H., Xie, L., Wu, Q., Zhang, B., Gao, Y., Zheng, W., & Shu,, X. (2013). Fruit and vegetable intake and the risk of colorectal cancer: results from the Shanghai Men's Health Study. *Cancer Causes Control*, 24(11), 1935-1945. https://doi.org/ doi:10.1007/s10552-013-0268-z.

Wahinda, K., Padilla, C. M., Denis, Z. N., Oliver, B.,Geraldine, L., Philippe, Q.,..& Severine, D. (2016). A conceptual framework for assessing of cumulative

exposure to air pollution at a fine spatial scale. *International Journal of Environmental Research Public Kealth, 13(*3), 319. https://doi.org/10.3390/ijerph13030319.

Wakefield J. (2007). Ecological studies revisited. *Annual Review of Public Health, 29*, 79-90. https://doi.org/10.1146/annurev.publhealth.29.020907.090821

Wong, C. M., Tsang, H., Lai, H. K., Thomas, G. N., Lam, K. B., Chan, K. P., Zheng, Q., Ayres, J. G., Lee, S. Y., Lam, T. H., & Thack, T. Q. (2016). Cancer mortality risks from long-term exposure to ambient fine particle. *Cancer Epidemiology Biomarkers Prevention, 25*(5):839–845. https://doi.org/10.1158/1055-9965.EPI-15-0626

World Bank (2013). Income Gini Coefficient. World development human indicators. thtp://data.worldbank.org/data-catalog/world-development-indicators

World Cancer Research Fund. (2018). *Diet, Nutrition, Physical activity, and Cancer: A global perspective.* Continuous update expert report 2018. Retrieved June 28, 2020. https://www.wcrf.org/sites/default/files/Summary-of-Third-Expert-Report-2018.pdf

Xu, Z., Wu, N., Zhou, H., Li, H., Yu, W., Fu, Z.,.., & Yin, P. (2017). Incidence of cancer in Shenzhen Guangdong province during 2001-2015. A prospective population-based study. *International Journal of Environmental Health and Public Health, 14* (1137), 2-16. https://doi.org/10.3390/ijerph14101137

Yamashita, S., Brudvik, K. W., Kopetz, S. E., Maru, D., Clarke, C. N., Passot, G., Conrad, C., Chun, Y. S., Aloia, T. A., & Vauthey, J. N. (2018). Embryonic Origin

of Primary Colon Cancer Predicts Pathologic Response and Survival in Patients Undergoing Resection for Colon Cancer Liver Metastases. *Annals of surgery*, *267*(3), 514–520. https://doi.org/10.1097/SLA.0000000000002087.

Yang, C.Y., Cheng, B. H., Hsu, T. Y., Tsai, S.S., Hung, C. F., T.N. Wu, T. N. (2000). Female lung cancer mortality and sex ratios at birth near a petroleum refinery plant. *Environmental Research, 83* (1), 33-40. https://doi.org/10.1006/enrs.2000.4038

Yuan, T., Chengshen, Y., Shiel, R., Hung, S., H., Chen, C., & Chan, C. (2018). Increased cancers among redidents living in the neighbourhood of a petrochemical complex: A 12-year retrospective study. *International Journal of Hygiene and Environmental Studies, 22*1(2), 308-314. https://doi.org/10.1016/j.ijheh.2017.12.004

Zaitsu, M., Tateuchi, T., Kobayashi, Y., Kawachi, I. (2019). Light to moderate amount of lifetime alcohol consumption and risk of cancer in Japan. *Cancer 126*(5), 1031-1040. https://doi.org/10.1002/cncr.32590

Zhang, Q., Berger, F. G., Love, B., Banister, C. E., Murphy, E. A., & Hofseth, L. J. (2018). Maternal stress and early-onset of colorectal cancer. *Medical Hypothesis, 121*, 152-159. https://doi.org/10.1016/j.mehy.2018.09.035