SAJE South African Journal of Economics
ESSA The Economic Society Of South Africa (Est. 1925)
South African Journal of Economics Vol. 85:3 September 2017

PEER NETWORKS AND TOBACCO CONSUMPTION IN SOUTH AFRICA

ALFRED KECHIA MUKONG*

Abstract
This paper deepens the empirical analysis of peer networks by considering their effects on both smoking participation and smoking intensity in South Africa, a country where majority initiate smoking at adolescent age. Peer networks are key in determining the smoking behaviour of youths, but the magnitude of the effects is still debated, questioned and inconclusive. I used a control function approach, a two-stage least square and the fixed effect method to address the potential endogeneity of peer network. The results suggest positive and significant peer effects on smoking participation and smoking intensity. While the network effects are consistently positive and significant, the magnitude of the estimates varies across methodological approaches with the instrument variable estimates generally lower. Including older adults in the peer reference group increases the peer effect estimates. Finally, using clusters as an alternative measure of network size in wave one, I show that peer effects are independent on network size but rather on network quality. Relative to the results of this paper, previous literature has documented larger peer effects on the decision to smoke. The findings suggest that policies (excise tax) that directly affect the decision to smoke and the smoking intensity of the peer reference group are likely to affect own smoking behaviour.
JEL Classification: I10, I12, D12, C36
Keywords: Peer network, smoking behaviour, control function, instrumental variable

1. INTRODUCTION

Cigarette smoking, an avoidable risk factor associated with cancer and other related heart diseases, is one of the leading causes of preventable and premature deaths each year (Silles, 2015). Globally, over five million premature deaths in 2000, over six million in 2014 and

* Corresponding author: Post-doctoral Research Fellow, South African Labour and Development Research Unit (SAL-DRU), School of Economics, University of Cape Town, South Africa. E-mail: alfred.mukong@uct.ac.za
I appreciate funding of the research by the Economics of Tobacco Control Project hosted by the South African Labour and Development Research Unit (SALDRU) at the School of Economics, University of Cape Town. I am indebted to the South African Journal of Economics (SAJE) anonymous referees who reviewed the paper and provided comments and suggestions that helped in shaping and improving the overall quality of the paper. I acknowledge the comments and suggestions from Grieve Chelwa that helped the evolution of this study from its inception to completion. I would also like to thank Corne Van Walbeek, Hana Ross and Justine Burns for their constructive criticisms. I am also grateful for the valuable comments and suggestions received from the editor, Jan Van Heerden and the anonymous reviewer of the Economic Research Southern Africa (ERSA) working paper series and from the Centre for the Study of African Economics (CSAE) conference on March 2016.
[Correction added on 10 August 2017, after online publication: The Acknowledgement section was previously omitted and has been added in this current version.]

 doi: 10.1111/saje.12166

an anticipated eight million by 2030 are smoking-related (WHO, 2015). A policy option that can help reduce future smoking-related deaths should therefore focus on reducing the prevalence of smoking, especially among youths, since adolescent smoking is a strong predictor of smoking addiction. A considerable body of research has shown that peers and peer relationship is a primary factor that influence cigarette smoking (Norton *et al.*, 1998; Powell *et al.*, 2005). While empirical studies have consistently provided evidence of significant peer effects on smoking decisions, the magnitude of peer influence on smoking decisions and alcohol consumption is still debated, questioned and not yet conclusive (Fowler and Christakis, 2008). This follows from the three interpretations of peer effects offered in Manski (1993), namely, the endogenous effects, exogenous effects and correlated effects.[1] With these identification challenges, recent studies have tried to purge the biases from peer effect estimates (Fletcher, 2010; McVicar and Polanski, 2014).

Some of these studies find positive and significant peer effects (McVicar and Polanski, 2014), but others argue that peer effects of substance use are weaker than identified in previous studies (Krauth, 2007; Duarte *et al.*, 2014) or even insignificant (Soetevent and Kooreman, 2007). According to McVicar and Polanski (2014), while such research have used numerous econometric techniques to provide evidence of peer pressure on cigarette smoking, focus has been on a limited reference groups. For instance, studies of peer effects in adolescent tobacco use rely on readily available school-based survey data, and uses the school, school grade or class as the reference group. The question is whether or not there are alternative measures of peer networks, as these studies are silent on the behaviour that takes place outside the school environment (non-school peers).[2] The identified peer effects from school-based survey data does not allow us to make generalisations of peer effects at national level, and hence national policies to reduce peer influence on smoking are made difficult. In addition, there is no evidence of peer influence on the intensity of smoking (the average number of cigarettes smoked by smokers) and what happens to adolescent peer effects when adults (above 24 years) are considered as part of their reference group. Finally, there is little evidence of peer effects on smoking in the context of developing countries, especially in Sub-Saharan Africa, where there is limited survey data on people's smoking behaviour.

In this paper, I deepen the empirical analysis of peer effects on cigarette smoking decisions by considering their effects on both the decision to smoke and on smoking intensity in South Africa. Building on the existing findings, I extend my analysis by introducing an alternative approach for measuring peer networks, and using a national

[1] In practice, it is empirically difficult to verify the effects peers exert on each other's substance use behaviour (Eisenberg, 2004). The difficulty stems from the problem of separating the impact of peer behaviour on own behaviour (endogenous effects), from the impact of peer characteristics (contextual or exogenous effects) and/or correlated unobservable factors (correlated effects) on own behaviour (Manski, 1993).

[2] Peer effects are externalities that occur when the action of a reference group affect the behaviour of others (McVicar and Polanski, 2014). Such effects have been studied in the context of labour market decisions (Oreopoulos, 2003; Burns *et al.*, 2010), education (Parker, 2012; Vardardottir, 2013; Chou *et al.*, 2015), welfare participation (Bertrand *et al.*, 2000; Dahl *et al.*, 2014), health outcomes (Deri, 2005; Kwon and Jun, 2015; Mukong and Burns, 2015) and smoking habit (Powell *et al.*, 2005; Nakajima, 2007; McVicar and Polanski, 2014; Duarte *et al.*, 2014). In the context of smoking habit, the considered reference group exclude individual contacts outside the school environment (mostly the community that might have had greater influence on smoking behaviour) (McVicar and Polanski, 2014).

representative panel data that permit the use of a broader and a more relevant reference group. The main focus is on individuals aged between 15 and 24 years. To check for sensitivity of the results, older adults are included. Cultural differences between countries may determine the extent to which smoking behaviour is influenced by peers (Gibbons *et al.*, 1995). For instance, a study in the Netherlands finds no peer effects (see Soetevent and Kooreman 2007), while studies from the United States and other European countries have demonstrated large, significant and positive peer effects on smoking behaviour of youths (see Powell *et al.*, 2005; McVicar and Polanski 2014). The question is, to what extent do the differences in peer effects reflect differences in methods, and to what extent does it reflect actual differences in peer effects across countries? Country specific case studies are essential, since the extent to which peer effect estimates for one country can be generalised to other countries has not been established and the magnitudes are still questionable and debatable.

If a peer can influence the smoking decision of others, interventions that reduce that peer's propensity to smoke will spread to his/her peers (Ali and Dwyer, 2009). However, robustly and accurately identifying peer effects estimates for policy intervention on smoking-related behaviour requires disentangling peer influence from spurious unobserved factors associated with peer selection. According to Fletcher (2010), policies that take advantage of peer effects may only achieve the desired objective if the common underlying attributes of the reference group drive behaviour more than the peer influence. The measure of peer in this paper is drawn not only from proximity in terms of geography, as has been the norm in the literature, but also from individuals who speak the same home language (ethnicity). This allows for the identification of the differences in the effects that could be exerted by different compositions of the reference groups (like cultural differences). I use a fixed effect and a control function (CF) approach to purge the potential biases emanating from peer effect estimates. This approach allows for a simple test of endogeneity of peer network and is more robust than the two-stage least square (2SLS) when instrumenting in a binary choice models.

The remainder of the paper is organised as follows: Section 2 describes the relevant institutions. Section 3 reviews theoretical insights of peer networks. Section 4 presents the data and descriptive statistics while Section 5 describes the empirical strategy. Section 6 presents the empirical results and Section 7 concludes the paper.

2. THE INSTITUTIONAL CONTEXT

While reduction in smoking prevalence may be regarded as a short-term goal for a tobacco control policy, the long-term benefit is an improvement in public health outcomes. In South Africa, smoking is still a significant problem affecting health. In 2013, the Cancer Society of South Africa reported that 44,000 of all deaths each year are from tobacco-related diseases. This is with inspite of the drastic decline in cigarette smoking that emanates from legislative steps and tax/price increases that aim to discourage tobacco consumption since the democratic transition in 1994. While there is a significant decline in tobacco consumption, the smoking rates varies significantly across population groups, gender, age cohorts, regions and income groups (Groenewald *et al.*, 2007).

Before the 1990s, tobacco tax increases was the main focus of government policy intervention (Asare, 2009) and since the early 1990s, there have been extensive regulatory reforms concerning tobacco consumption. This includes an increase in excise tax, limits on public smoking and strict control over advertising (Van Walbeek, 2004; Boshoff, 2008). For instance, the Minister of Health in the early 1990s was given power to restrict

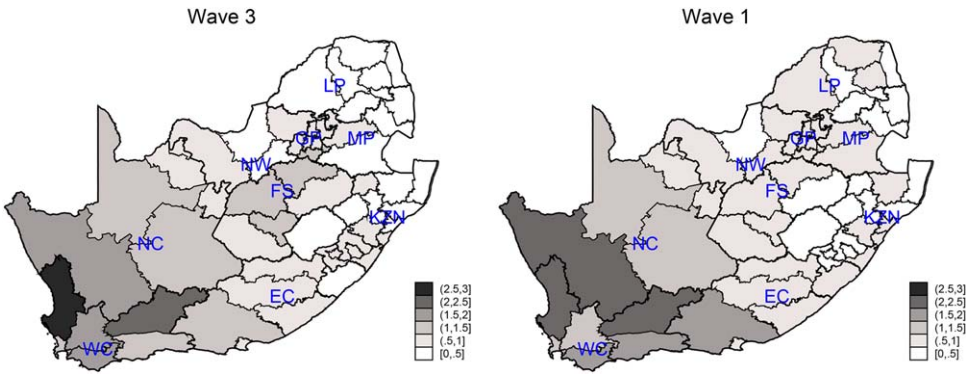

Figure 1. District smoking rates as a proportion of the national smoking rate [Colour figure can be viewed at wileyonlinelibrary.com]
Notes: Colouring goes from dark, denoting high smoking rate, to light denoting low smoking rate relative to the national average. WC: Western Cape; EC: Eastern Cape; NC: Northern Cape; FS: Free State; NW: North West; KZN: KwaZulu-Natal; GP: Gauteng; MP: Mpumalanga; and LP: Limpopo.

smoking in certain public places, to illegalise the selling of cigarettes to children under the age of 16 years and to force cigarette advertising to carry health warnings (Leaver, 2002). The largest of these policies was the Tobacco Products Control Amendment Act (TPCAA) in 1999. The 1999 Tobacco Products Control Act (TPCA) amended the 1993 TPCA by prohibiting the advertisement and promotion of tobacco products, the free distribution of tobacco products and the receipt of gift or cash prices in contests, lotteries or games to the buyer of tobacco products. The existence of these policies sparked research on the economics of tobacco in South Africa, with particular focus on price sensitivity (Van Walbeek, 1996, 2000; Abedian and Jacobs, 2001).

To keep to the recent requirements of the World Health Organisation Framework Convention on Tobacco Control (WHO FCTC), the government has further strengthened its tobacco control policies by introducing new and non-tax policies in 2007 and 2008. The policies include, an increase in smoking fines, illegalisation to smoke in a car with children under the age of 12 years and warning pictures on cigarette packs (Government, 2007). While there is evidence of how price changes among others explains smoking behaviour in South Africa, existing cultural differences across region is a call for concern for more local studies to identify the effects of peer network on tobacco consumption in South Africa. In addition, the fact that over 87% of smokers in South Africa started smoking before age 25 years, an age range characterised by high degree of peer influence, necessitate the need to understand the role of peer networks on individual smoking behaviour.

The study of peer networks on smoking behaviour in South Africa is further motivated by several interesting facts. First, the smoking prevalence levels and trends by demographic characteristics (age, race and gender) and geography (province) has been consistently different since the early 1990s. For instance, since 1993, the Western Cape and the Northern Cape has recorded the highest level of cigarette consumption with Limpopo and Mpumalanga having the lowest rates (van Walbeek, 2002). See Fig. 1 for recent evidence from our data. In addition, the highest smoking prevalence is found to be in more affluent provinces and those with relatively high population of coloured

people. Finally, the majority of smokers initiate smoking at the adolescent age, suggesting the likelihood of peer network effects.[3] Rather than rely on intuitive assumptions or on evidence from other countries with different cultural and demographic settings, this paper attempts to tease out these effects in the context of South Africa.

3. THEORETICAL PERSPECTIVES

There are several theoretical frameworks that explain the process through which social interaction influence an individual's risky behaviour, especially in terms of alcohol consumption, drug and tobacco use. A number of these perspectives include social learning theory (Akers, 1977; Simpson, 2000), the social identity theory (Abrams and Hogg, 1990), primary socialisation theory (Oetting and Donnermeyer, 1998), social network theory (Granovetter, 1973; Wasserman and Faust, 1994), the social bonding theory (Hirschi, 1969), a general theory of reasoned action (Fishbein and Ajzen, 1975), peer cluster theory (Oetting and Beauvais, 1986), the triadic theory of influence (Ajzen, 1985) and the social development theory (Hawkins and Weis, 1985). This paper draws on the social learning, primary socialisation, social identity and the social network theories. These theories explain social processes, such as friend selection, interpersonal influence and behaviour imitation and provide unique insights in understanding tobacco use effects of peer network.

In the social learning theory, cognitive mediation is considered essential in the acquisition and maintenance of smoking behaviour (Akers, 1977; Simpson, 2000). In this perspective, behaviours are learned by observing others involved in a similar behaviour. Here, the direct influence of parents and peers are considered as the primary social factors, while the media and indirect reference groups are regarded as secondary social factors. More intimate relationships that occur in youths' early experiences are crucial in their social learning process than those that come later in their lives. In terms of tobacco consumption, youths are regarded as being most likely to imitate smoking behaviour of their close contacts. The theory, therefore, predicts that social learning on substance use can progress to frequent or sustained patterns, to the extent that even negative sanctions and unfavourable definitions of tobacco, such as the negative health consequences may not offset the decision to smoke.

The primary socialisation theory is a reformulation of peer cluster theory of drug initiation (Oetting and Beauvais, 1986; Oetting and Donnermeyer, 1998). The theory identifies that the family and peer clusters are the primary contexts through which norms and behaviours are learned (the primary focus of this study). Because the media and local institutions influence families and peer clusters, they are regarded to have an indirect influence on norms and behaviours. It underscores that rational bonds between individuals, their family as well as peers are important in transmitting information about norms and behaviours. One argument is that individuals are unlikely to engage in substance use (drugs, alcohol and tobacco), if the bond between them and their families are pro-social and strong (Hirschi, 1969). On the other hand, the influence of peer cluster is heightened if the bond between individuals and their families are weak, especially if the cluster promote substance use. In this regard, peers are considered a main source of substance use.

The social identity theory focuses on an individual's self-concept as a group member and distinct social groups (Abrams and Hogg, 1990). For instance, in the context of self-

[3] According to social learning theory youths are most likely to imitate smoking behaviour of their close contacts, and direct influence of parents and peers are the primary factor for social learning.

concept, individual characteristics matters, whereas in the social categorisation, the characteristics of the group play an important role. Individuals are expected to act according to their personal norms if their personal identity is significant, but act in accordance with the group, if the social identity is important. The theory does not consider the similarities among group members as a source of social pressure, but rather assumes that members adopt those norms and behaviours central to the group. For example, it considers that the smoking habits of members of the group are likely to be similar if smoking status is central to the social identity of the group.

The social network theory builds on the interdependence between individuals and the existing rationale between individuals in a social system or a targeted population identified by specific boundaries, such as a school, classroom and neighbourhood setting (Wasserman and Faust, 1994). This is the approach use in this paper to identify the size of one's network, since individual's actual contact cannot be determined with certainty. The theory assumes that individuals interact with each other and serve as a significant reference point in each other's decision making, leading to transfer of information and resources. The attitude, perception and behaviour of an individual in the network are influenced by his/her location and pattern of relations with others in the network. The theory has been used to examine the way smoking norms might be communicated within and transmitted across the system. It considers two types of individuals, those central (highly connected) and those marginal (loosely connected) to the system (Kobus, 2003). While the former is more likely to adopt non-controversial issues, the latter is more likely to adopt controversial issues such as smoking (Granovetter, 1973). The theory suggests the need to consider a larger social system in understanding peer network on tobacco use (Kobus, 2003).

Each of these theories provides a framework for understanding social processes and youths' decision to engage in risky behaviour like cigarette smoking. The theories differ in specific social and cognitive processes they present, but they all place importance on the type of peers with whom individuals interact. Explicitly, considering other factors, each of these theories suggests directly or indirectly that the norms and behaviour of an individual's (especially teenagers) peers are imperative in determining behaviour. That is, teenagers are more likely to smoke if their peers smoke and reinforce smoking behaviour, but less likely to smoke if their primary contacts (the family) are non-smokers. Each theory provides a unique contribution to the understanding of peer network on individual behaviour. While social learning theory highlights mechanism of social influence, the primary socialisation theory points to the importance of individual characteristics and rational bonds between individuals and their family and peers. The social identity theory points to group comparison and adoption of social identity, and social network theory highlights the importance of location in the system and pathway of information exchange. Most studies on peer network do not specify the theoretical perspectives guiding the research and the assumptions for selecting variables. These theoretical perspectives, when woven together, provide a more comprehensive framework for studying peer network on cigarette smoking. They give a clearer picture on the aspects of peer influence, and when and how this influence affect individual as well as the smoking behaviour of the group.

4. DATA AND DESCRIPTIVE STATISTICS

The analysis is based on the National Income Dynamic Study (NIDS), which is the first nationwide set of panel survey data designed to track changes in the well-being of South

Africans over time. The data provides information on a representative sample of households and their members living across the country. It combines household-level interviews with questionnaires addressed to both adults (aged 15 or older) and children in the household. In this paper, I only consider information of adult household members, since there is no information on the smoking behaviour of teenagers. Currently, there are three waves available. The first wave, completed in 2008, provides information on 7,236 households with 16,781 adult individuals. The second wave was conducted in 2010 on 9,734 households with 21,880 adults. The third wave, carried out in 2012, provides information on 10,236 households with 22,481 adult individuals. All of the surveys collect detailed information on household and individual demographic and socioeconomic characteristics.

The adult questionnaires collected information on individual smoking behaviour. This includes whether or not the individual smokes cigarettes and whether or not he/she ever smoked cigarettes regularly. Both smokers and ex-smokers were asked the age at which they first smoked cigarettes and only ex-smokers were asked when they last smoked cigarettes regularly. Finally, individuals were asked to indicate on average the number of cigarettes they smoke per day. In this analysis, I identify peer effects only on two of these questions, namely, the individual's decision to smoke and the average number smoked daily. For peer effects on the decision to smoke, I limit the sample to all adults between the ages 15 and 24. However, I further control for older adults (at most aged 45) to demonstrate how their inclusion affects the peer effect estimates. Surveys weights are used to correct for any imbalances between sample characteristics and known population parameters.

There are two dependent variables, smoking participation and smoking intensity. For the decision to smoke, the dependent variable is a 0–1 decision to smoke, which is the declaration of each survey respondent whether or not he/she smokes. I then disaggregate the dependent variable to identify the effects of peer networks on an individual's decision to smoke. The second dependent variable is the average number of cigarettes smoked daily. This is a continuous variable and is restricted only to current smokers. The intensity to smoke variable is, therefore, the logarithm of the number of cigarettes an individual smoke daily.

In this paper, I test for the presence of peer network effects on cigarette smoking decisions by constructing a variable that takes into consideration the quality and size (quantity) of the peer network. The quantity of peer network captures the fact that the larger the number of people who live in close proximity and speak the same home language, the larger the available contacts, that is, people that may influence one's smoking behaviour.[4] The quality of peer network captures characteristics such as cultural differences in beliefs about smoking. Contacts drawn from high cigarette smoking groups are more likely to have a stronger influence on the decision to smoke and smoking intensity. The smoking behaviour (the relative proportion of smokers or general sentiment towards smoking) of a language group provides a measure of peer network quality.[5]

[4] Evidence suggests that those whose home language is not English in the United States interact mainly with others from their language group (Alba, 1990). In the study of American born White ethnics, Alba (1990) used mother tongue as a determinant of ethnic identity and showed that half of all non-related childhood friends belonged to the same ethnic groups.

[5] Language group refers to all individuals in South Africa who speak the same home language.

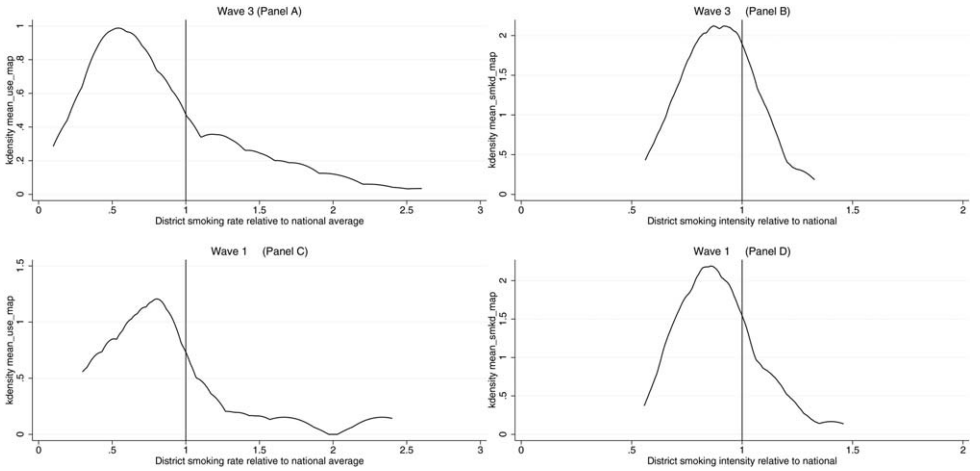

Figure 2. South Africa's Regional Smoking Rate/Intensity in wave 1 and wave 3
Note: Values of 1, (>1), (<1) on the horizontal axis denote districts with a smoking rate/intensity equal to, larger than and/or smaller than the South African smoking rate, respectively.

In Fig. 1, I present a preliminary look at the spatial aspect of the distribution of regional or district smoking rates in South Africa. This means that one, less than one and/or greater than one denotes a district with a smoking rate equal to, smaller than, and/or larger than the national smoking rate, respectively. The figure immediately indicates that South Africa is characterised by few districts (or provinces) that have smoking rates above the national level, and relatively many districts have smoking rates below the national level (see Panel A of Fig. 2).

In terms of spatial distribution of smoking rate across districts, I find little difference between wave 1 and wave 3 (see Fig. 1 and Panel C of Fig. 2 for comparison). In addition, the smoking rate in some of the regions are more than two times higher than the average smoking rate in South Africa. Specifically, the darker the colour a region is on the map, the higher the smoking rate relative to the national smoking rate. The map depicts that Western Cape has the highest proportion of smokers relative to the national average. This is immediately followed by Northern Cape, Free State and Gauteng. Limpopo and Mpumalanga have the lowest rates. This is consistent with the 1994 and 2002 statistics presented in van Walbeek (2002).

Figure 3 and Panel B of Fig. 2 indicate that the number of districts with smoking intensity above the national average are almost evenly distributed in relation to those with averages below the national average. These findings are consistent across the waves (see Fig. 3 and Panel D of Fig. 2 for comparison). As is the case with smoking rate, the darker the colour a region is on the map, the higher the smoking intensity relative to the national smoking intensity.

Table 1 reports summary statistics for the sample by individual smoking behaviour, revealing the interesting differences between smokers and non-smokers. The results demonstrate that only 20% of the sample are current smokers and 80% are non-smokers. In terms of smoking intensity, an average smoker in South Africa smokes nine cigarettes a day. The Coloured and White population have a higher percentage of smokers and a

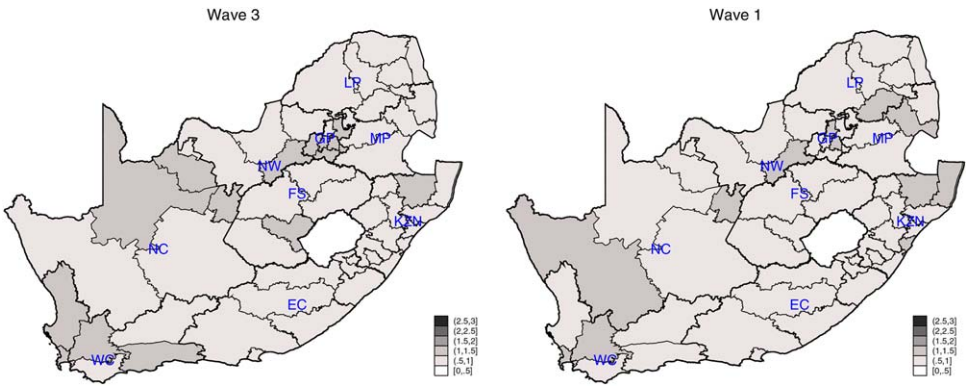

Figure 3. District smoking intensity as a proportion of the national smoking intensity [Colour figure can be viewed at wileyonlinelibrary.com]
Notes: Colouring goes from dark, denoting high smoking intensity, to light denoting low smoking intensity relative to the national average. WC: Western Cape; EC: Eastern Cape; NC: Northern Cape; FS: Free State; NW: North West; KZN: KwaZulu-Natal; GP: Gauteng; MP: Mpumalanga; and LP: Limpopo.

lower percentage of non-smokers relative to their share in the overall population. In contrast, Africans (Blacks) and Indians have a lower percentage of smokers and a higher percentage of non-smokers relative to their share in the population as a whole. For instance, 10% of the overall sample are White, but over 13% of the sample of smokers are White and the proportion of Coloured in the sample of smokers is twice their proportion in the whole population. In addition, over 65% of the sample of smokers and 83% of the sample of non-smokers are Africans, relative to their share of the overall population of 78%. The average age of individuals in the overall sample and that of non-smokers is 37 years relative to 39 years for the sample of smokers.

The proportion of women in the overall sample (52%) is more than double their proportion in the sample of smokers (22%) but less, relative to the sample of non-smokers (62%). On the contrary, the percentage of men in the sample of smokers (78%) is greater compared to their share in the sample (48%). Individuals who drink most often have a higher proportion of those who smoke (31%) and a lower percentage of those who do not smoke (5%) relative to their share in the entire sample (10%). Individuals with at most some secondary education have a higher proportion of those who smoke (84%) and a lower proportion of non-smokers (78%) relative to their share in the sample (79%). The converse holds true for those with some university education.

While non-religious individuals have a larger proportion of smokers (29%) relative to their share in the sample (18%), Christians have a relatively low proportion of smokers (69%) compared to their share in the entire sample (80%). Another interesting difference between smokers and non-smokers lies in their ethnicity (measured by language spoken at home). While those whose home language is Afrikaans, English and Sesotho have a higher proportion of smokers relative to their share of the population, IsiTsonga, Tshivenda, Siswati, Setswana, Sepedi, IsiZulu and IsiXhosa speakers have a lower percentage of smokers relative to the share in the entire sample.

Table 1. Mean statistics for sample by smoking behaviour in wave 3

Variable	All		Smokers		RPS	Non-smokers		RPNS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Individual does not smoke	0.80	(0.40)						
Individual smokes	0.20	(0.40)						
Average number of cigarette smoked a day	8.53	(6.93)						
Average age smokers first smoked cigarette			20.74	(7.15)				
Age	36.69	(16.17)	38.72	(13.59)	1.06	36.61	(16.66)	1.00
Individual is African	0.78	(0.42)	0.65	(0.48)	0.83	0.83	(0.38)	1.06
Individual is Coloured	0.09	(0.29)	0.20	(0.40)	2.22	0.06	(0.24)	0.67
Individual is Indian	0.03	(0.16)	0.02	(0.15)	0.67	0.02	(0.16)	0.67
Individual is White	0.10	(0.30)	0.13	(0.34)	1.30	0.08	(0.28)	0.80
Individual is female	0.52	(0.50)	0.22	(0.41)	0.42	0.62	(0.48)	1.19
Individual is male	0.48	(0.50)	0.78	(0.41)	1.63	0.38	(0.48)	0.79
Individual does drink often	0.90	(0.30)	0.69	(0.46)	0.77	0.95	(0.22)	1.06
Individual drinks often	0.10	(0.30)	0.31	(0.46)	3.10	0.05	(0.22)	0.50
Individual has no formal education	0.06	(0.24)	0.06	(0.23)	1.00	0.07	(0.25)	1.17
Individual has at most Metric	0.79	(0.41)	0.84	(0.37)	1.06	0.78	(0.42)	0.99
Individual has university education	0.15	(0.35)	0.10	(0.30)	0.67	0.16	(0.37)	1.07
Home language is IsiNdebele	0.01	(0.12)	0.01	(0.10)	1.00	0.01	(0.12)	1.00
Home language is IsiXhosa	0.18	(0.38)	0.16	(0.37)	0.89	0.19	(0.39)	1.06
Home language is IsiZulu	0.25	(0.43)	0.20	(0.40)	0.80	0.25	(0.43)	1.00
Home language is Sepedi	0.11	(0.31)	0.08	(0.26)	0.73	0.13	(0.33)	1.18
Home language is Sesotho	0.09	(0.28)	0.12	(0.33)	1.33	0.08	(0.28)	0.89
Home language is Setswana	0.10	(0.30)	0.08	(0.26)	0.80	0.11	(0.32)	1.10
Home language is Siswati	0.02	(0.15)	0.01	(0.12)	0.50	0.03	(0.16)	1.50
Home language is Tshivenda	0.02	(0.13)	0.01	(0.08)	0.50	0.02	(0.14)	1.00
Home language is IsiTsonga	0.03	(0.18)	0.02	(0.13)	0.67	0.04	(0.19)	1.33
Home language is Afrikaans	0.12	(0.33)	0.23	(0.42)	1.92	0.09	(0.28)	0.75
Home language is English	0.06	(0.25)	0.08	(0.27)	1.33	0.05	(0.22)	0.83
Peer network	0.48	(1.11)	1.42	(2.25)		0.26	(0.24)	
Individual is non-religious	0.18	(0.38)	0.29	(0.45)	1.61	0.15	(0.36)	0.83
Individual is Christian	0.80	(0.40)	0.69	(0.46)	0.86	0.83	(0.38)	1.04
Individual is Muslim	0.01	(0.10)	0.01	(0.08)	1.00	0.01	(0.10)	1.00
Individual is Jewish/Hindu	0.01	(0.12)	0.02	(0.14)	2.00	0.01	(0.12)	1.00

Notes: Standard deviation in parentheses. The sample includes all individuals between the age 15–45 years. There are two dependent variables, namely, individual is or is not a smoker and the average number of cigarettes smoked per day. Column (1) presents statistics for the entire sample of individuals; columns (3)-(5) present statistics for smokers only and columns (6)-(8) present statistics for non-smoker only. RPS = relative proportion of the smoking sample as a ratio of the overall sample and RPNS = relative proportion of the non-smoking sample as a ratio of the overall sample. If RPS or RPNS is greater than one, it signifies that higher percentage of smokers or non-smokers relative to the share of smokers and non-smoker in the overall population and vice versa.

5. EMPIRICAL STRATEGY

The econometric model for the effects of peer network on the decision to smoke and the smoking intensity of an individual i from language-district group j at time t (time refers to the different waves) is written as:

$$Y_{ijkt}=\beta_0+\beta_1 N_{ijkt}+\beta_2 X_{ijt}+\beta_3 P_{ijt}+\beta_4 L_{jt}+\beta_5 G_{jt}+V_{jkt}\gamma+\varepsilon_{ijt} \quad (1)$$

Where N_{ijkt} is a measure of peer network, which is the product of quantity of contacts (network size) (V_{jkt}) and quality of network ($\overline{Y_{kt}}$). X_{ijt} represents individual characteristics and P_{ijt} are household and/or parent characteristics. L_{jt} is a language dummy that control for unobservable language (ethnic) group specific characteristics common to all individuals having the same home language (such as ethnic attitude towards cigarette smoking),

G_{jt} is a geography dummy that controls for district specific characteristics common to all individuals within the same district (such as easy access to cigarettes), and ε_{ijt} is the random error term. Y_{ijt} is a binary outcome variable when modelling the decision to smoke and the logarithm of number of cigarettes smoked daily in the case of smoking intensity and $\overline{Y_{kt}}$ is the average number of smokers in each language group.

The estimation of peer network models is generally made difficult, since data on one's actual contacts and the extent of one's peer are rarely collected. In many instances, mean neighbourhood attributes are used to identify one's network, with the assertion that contacts are randomly distributed within the neighbourhood. While international evidence suggests several dimensions through which networks can be operationalised, (along religious, racial, ethnic, age and education lines for example), I used the language (ethnic) and geography to measure network, since individuals speaking a common language and residing in the same community are more likely to spend time together and learn from each other. Although other network dimensions remain important, most international evidence, especially in the United State suggest the use of language grouping (see Alba, 1990; Bertrand *et al.* 2000). In the context of South Africa, Burns *et al.* (2010) argued that speaking a common language was an important source of information and used language and district grouping as a measure of network to identify the impact of social networks on employment in South Africa. As a robust check, I mimic the language grouping dimension by considering educational attainment as an alternative dimension along which peer network may operate.

Network size (quantity of contacts) is measured as $\ln\left(\frac{V_{jkt}/Ajt}{L_{kt}/T_t}\right)$, where V_{jkt} measures the number of individuals in district j belonging to language group k at time t; A_{jt} is the number of individuals in district j; L_{kt} is total number of individuals in the sample that belong to the same language group; and T_t is the total sample used in our analysis. It is the case that small districts or language groups will have small available contacts even if there is full concentration in such districts and within this language groups. Using proportions resolve the problem of under-weighting of small districts as well as small language groups (Bertrand *et al.*, 2000). For the quality of network, the smoking rate by language group as a ratio of the smoking rate in the entire sample is used. Precisely, it is measured as the mean deviation of the group's level of smoking relative to the smoking prevalence rate of the entire sample. Most literature on peer effects on smoking assume the mean smoking rate for each peer group k excluding the individual i as a measure of network quality. Bertrand *et al.* (2000) argue that this approach introduces bias as it may reflect the unobserved characteristics the individual has in common with others in the group. They propose the use of relative means. This approach differs from the standard approaches used in the tobacco literature, but has been used significantly in the welfare participation (Bertrand *et al.*, 2000), health care utilisation (Deri, 2005) and labour market decisions (Burns *et al.*, 2010) literature.

The main focus of this paper is on the endogenous effects β_1, which explains the extent of peer network on the smoking decision and smoking intensity of individuals. Positive and significant estimates of peer network (β_1) indicates that any policy that influence an individual's smoking decision within a reference group will to some extent affect the smoking behaviour of others in the network (Fletcher, 2010). However, the estimated peer effects (β_1) is likely to be biased, if the reflection problem pointed out by Manski (1993) is not properly controlled. The inclusion of the language group, location

(district) and interacted language and district fixed effects purge some of the bias resulting from the correlated unobservable characteristics (correlated effects).

To deal with biases resulting from any omitted individual characteristics correlated with V_{jkt} (network size), I include V_{jkt} as a control in the specification. Since the language group fixed effects L_{jt} are included in the estimation, I therefore exclude the direct effects $\overline{Y_{kt}}$ (network quality). The remaining potential bias not accounted for in this specification results from Manski (1993) reflection problem. The reflection problem emanates from the fact that the individual himself can affect the behaviour of his/her peers and at the same time the behaviour of his/her peers influence his own behaviour (source of endogeneity). This poses an identification threat which according to Manski (1993) the true β_1 can only be identified with the use of an instrumental variable approach.

5.1 Identification Strategy

I address the reflection problem by using a control function approach for a dichotomous dependent variable to provide a causal interpretation to β_1. In the control function approach, the endogenous variable is regressed on the instrumental variable(s) and the other explanatory variables and the residuals are saved. In the second step, a probit model for an individual's smoking decision is estimated as a function of the endogenous variable, the exogenous variables and the residuals (Wooldridge, 2010).[6] This approach is similar to the two-stage least square (2SLS) but allows us to test whether or not the network variable is actually endogenous and it provides consistent estimates (Rivers and Vuong, 1988). However, this hinges on the assumption that the instruments are exogenous or valid. The challenge identified in most of the literature is getting variables that are correlated with the peer network and has no direct effect on the individual's decision to smoke. I expect that the smoking behaviour of the parents of the peers will directly affects peers' smoking behaviour, but not the individual's own smoking behaviour. Balia and Jones (2008) showed that parent smoking behaviour has a significant influence on individual lifestyle choices. In addition, Powell *et al.* (2005) suggested that peer network effects are robust to a set of instruments that draw from measures of peer parent characteristics such as marital status, education level and parent–child discussion level.

A potential criticism of this instrument could be that omitted (unobservable) variables such as cigarette price, advertising and cultural norms. might be correlated with parental smoking, peer smoking and individual smoking behaviour. Angrist (2014) highlighted that empirical evidence that control for peer characteristics that are related with individual characteristics provide the most compelling results on the nature of peer effects. Excluding unobservable factors that might be correlated with parental smoking, peer smoking and individual smoking behaviour renders the IV approach inadequate in addressing the reflection problem. I minimise this bias by including language, area and the interacted language and area fixed effects. In addition, tobacco policies in South

[6] The control function approach has some advantages over other nonlinear two-step approaches that appear to mimic the 2SLS estimation of the linear model. Unlike the control function approach, getting appropriate standard errors is difficult and simply inserting fitted values of the endogenous variable from a 2SLS does not provide a formal test for existence of endogeneity. Estimates from the fitted values approach (2SLS) are not consistent and adding other functional forms of the endogenous variable is cumbersome and prone to mistakes as the fitted values are strictly limited to the structural equation (Wooldridge, 2010).

Africa prohibit the advertisement and promotion of tobacco products, and cigarette prices do not vary across individuals or regions. The omission of these variables do not limit, nor affect the strength of the instrument. I test for existence of endogeneity of the peer network measure using the Rivers and Vuong (1988) endogeneity test. According to Wooldridge (2010), it makes sense to compare the 2SLS estimates of a linear probability model (LPM) with the average partial effects from the probit model with an endogenous variable. For this reason, I compare the 2SLS estimates of the LPM with or without fixed effects to the consistent estimates of the control function approach.

6. EMPIRICAL RESULTS

While peer smoking behaviour may affect individual (own) smoking decision, the reverse may be true. In this case, the coefficient on peer networks will be biased and the effects of peer networks will be measured incorrectly (Angrist, 2014). I use instrumental variables (IV) regression (Wooldridge, 2010) to account for the endogeneity of peer networks. IVs that are good predictors of peer networks but are not independently correlated with the outcome are used to purge the bias. A control function, a 2SLS, and a fixed effect 2SLS approach is used to investigate the endogeneity of peer network (see Table 3). At the first stage I regress peer networks against the IV and other covariates using ordinary least square (OLS) to ensure the IV is a strong predictor of peer networks (see Table 9). It is evidence from all specifications that the instrument is a strong predictor of peer network. The predicted peer networks from the first stage is then included in the second stage.

Table 2 presents the peer effect estimates controlling for district and language group fixed effects.[7] The assumption underlying the results in this table is that peer effects on an individual's smoking behaviour are exogenous. The unboundedness of the estimated probabilities on the unit interval is considered a serious problem that may result in biased and inconsistent estimates of the LPM. The potential bias increases with the relative proportion of LPM predicted probabilities that fall outside the unit interval. Conversely, Horrace and Oaxaca (2006) argue that if few predicted probabilities lie outside the unit interval, the LPM is expected to be largely unbiased and consistent. In this case, the use of the LPM is not entirely problematic, since robust standard errors can commonly be used (see Paxton, 1999).

In this paper, I find that the proportion of LPM predicted probabilities that lie outside the unit interval ranges from 0.24 to 0.27% for all specifications. So it appears that according to Horrace and Oaxaca, the LPM estimates are unbiased and consistent. In this case, the LPM is preferred to the probit model, since the latter suffers computational difficulties in the presence of fixed effects (see Bertrand *et al.*, 2000). For sensitivity, I compare fixed effect estimates from both the LPM and the logit model (see Table 2 and Panel A of Table 5 for comparison). First, I consider individuals aged between 15 and 24

[7] Adding fixed effects to any binary outcome model (especially the probit) induces bias in the coefficient and standard errors (incidental parameter bias). In addition, it is near certainty that any probit estimation incorporating a nontrivial number of fixed effects will produce bias results (Baltagi, 2008). For the use of fixed effects in social sciences, there have been a switch from a standard normal probit to a logit model. The logit fixed effects are not dissimilar to multiple linear regression in that it filters out the fixed effects (Baltagi, 2008).

Table 2. Regression estimates of peer network as fixed effects are included (aged 15–24)

Variables	Wave 3				Wave 1			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Quantity of contacts	−0.05***	−0.13***	−0.21***	−0.22***	−0.05***	−0.08***	−0.07***	−0.09***
	(0.01)	(0.02)	(0.03)	(0.04)	(0.01)	(0.02)	(0.02)	(0.01)
Peer network	0.05***	0.06***	0.07***	0.12***	0.13***	0.15***	0.15***	0.22***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)	(0.04)	(0.04)	(0.01)
Indirect network effects	0.04***	0.04***	0.05***	0.10***	0.10***	0.11***	0.11***	0.17***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Age	0.08**	0.08**	0.08**	0.06	0.08***	0.08***	0.08***	0.07*
	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
Age squared	−0.00*	−0.00	−0.00*	−0.00	−0.00**	−0.00**	−0.00**	−0.00*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Individual is male	0.11***	0.10***	0.10***	0.10***	0.15***	0.15***	0.14***	0.08***
	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
Individual drinks alcohol often	0.32***	0.31***	0.30***	0.29***	0.24***	0.24***	0.25***	0.32***
	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	(0.07)
Constant	−0.88**	−0.70*	−0.09	0.03	−0.86***	−1.07***	−1.23***	−1.29***
	(0.37)	(0.37)	(0.37)	(0.38)	(0.32)	(0.30)	(0.31)	(0.35)
Observations	5,533	5,533	5,533	3,054	4,579	4,579	4,579	2,417
R-squared	0.28	0.32	0.35	0.47	0.35	0.38	0.39	0.51
District fixed effects	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Language group fixed effects	No	No	Yes	Yes	No	No	Yes	Yes

Notes: Standard errors are given in parentheses. I use Wave 1 and Wave 3 of NIDS, and restrict the sample to young adults (aged 15–24). Results from wave 3 are presented in columns (1), (2), (3) and (4), and wave 1 in column (5) to column (8). The basic sample includes all individuals who are between the ages 15 and 24, and who have one of the 11 languages in South Africa as his/her home language and available information on district of residence. In columns (1) and (5), all possible fixed effects are excluded. In columns (2) and (6) district fixed effects are included. In columns (3) and (7) both district and language fixed effects are included. In columns (4) and (8) all fixed effects and some parental characteristics are included. The dependent variable for all specifications is a dummy equal to 1 if the respondent smokes. Other controls include, race dummy, education dummy, religious dummy, parent education and smoking behaviour. The peer network variable is calculated as the product of quantity and quality of contacts. The quantity of contact is measured as $\ln\left(\frac{V_{jk}/Aj}{L_k/T}\right)$, where V_{jk} measures the number of individuals in district j belonging to language group k; A_j is the number of individuals in district j; L_k is total number of individuals in the sample that belonging to the same language group; and T is the total sample used in our analysis. It is the case that small districts or language groups will have small available contacts even if there is full concentration in such districts and within this language groups. Using proportions resolve the problem of underweighting of small districts as well as small language groups. For the quality of network, the smoking rate by language group as a ratio of the smoking rate in the entire sample is used. Precisely, it is measured as the mean deviation of the group's level of smoking relative to the entire sample. ***Statistically significant at 1% level; **statistically significant at 5% level and *statistically significant at 10% level.

and display their results from both waves in Table 2, taking account of parental education and smoking behaviour. Second, I consider individuals aged between 15 and 45 and the results from both waves are presented in Panel B of Table 5 for comparison.

It is interesting that the peer network effects on an individual's decision to smoke remain highly significant across waves, across different age groupings across different dimensions through which networks operate (see Table 10 for an alternative dimension) and after controlling for language group and location unobservable characteristics. The network effects for age group between 15 and 24 years ranges from 5 to 12% in wave 3,

and up to 22% when wave 1 is considered (see peer network variable in Table 2).[8] The network effects are higher when older adults (aged 25–45) are included as part of the reference group (see Panel B of Table 5). This indicates that the smoking behaviour of older adults is likely to play a significant role on the smoking behaviour of younger adults. In addition, the magnitude of peer effects increases with the inclusion of fixed effects. Ali and Dwyer (2009) and Fletcher (2012) also find larger peer effects in the presence of fixed effects. In general, the results suggest that peer networks increase the probability of own smoking. By implication, policies that target only the smoking behaviour of younger adults (15–24 years) have lower effects than those that cut across all adults (15–45 years). The negative effects of quantity of contacts indicates that the average number of peer who smoke and not just the size of peer network are responsible for the positive peer effects.

In terms of magnitude, these estimates are close to those obtained by McVicar and Polanski (2014), but far below those obtained in other studies (see Powell *et al.*, 2005; Ali and Dwyer, 2009; McVicar, 2011). The high magnitudes from these studies are not surprising, since they focus mainly on adolescent who are most likely to imitate the behaviour of their close contacts. In addition, identifying the estimates of peer effects on adolescent smoking across 26 European countries, McVicar (2011) showed that peer effects varies significantly across countries. Using longitudinal data to estimate peer network effects on adolescent smoking, Ali and Dwyer (2009) found a decline in peer effects 2 years after the first wave, irrespective of the measure of peer network used. My findings confirm to this, showing a huge decline in peer effects four years after the first wave.

Individual characteristics are also important in determining the probability of smoking. The probability of smoking increases with age until aged 25 and men are between 10 and 18 percentage more likely to smoke than women. Individuals who drink alcohol often are more likely to become smokers than those who do not, or drink occasionally. Individuals who are Coloured, Indian and White are more likely to smoke relative to their Black counterparts with the association highly significant among Coloureds than Whites and Indian. The results suggest that Christians and Muslims are significantly less likely to smoke than people with traditional or non-religious beliefs. These findings are similar to that of Ali and Dwyer (2009), who demonstrate that religious individuals have a lower probability of smoking than non-religious individuals. In general, the results of the LPM are similar to those of the logit model (see Table 2 Panel A of Table 5 for comparison).

In Table 3, I present the instrumental variables (IV) estimates derived from different specifications (the two-step control function approach using a probit model and the Two-Stage Least Square (2SLS)) with and without area fixed effects. The different specifications allow the results to be comparable with previous literature and provide a

[8] An increase in tobacco prices and the prevalence of tobacco control policies are likely to reduce the magnitudes of peer network (Powell *et al.*, 2005). They show that there is a potential for social multiplier effects if the literature on peer smoking effect takes into account the exogenous changes in cigarette taxes and tobacco control policies. Specifically, they find that the omission of these variables reduces peer networks by 0.06. With this evidence, the higher peer network estimates in wave 1 relative to wave 3 are not surprising given the tobacco control policies introduced between 2007 and 2008 and the price difference between 2008 and 2012.

Table 3. Marginal effects and regression estimates of peer network after controlling for endogeneity

Variables	Wave 3						Wave 1					
	Two-step CF		IV without FE (2SLS)		IV with FE (2SLS)		Two-step CF		IV without FE (2SLS)		IV with FE (2SLS)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	11	12
Quantity of contacts	−0.05***	−0.05***	−0.07***	−0.04***	−0.07***	−0.07***	−0.04***	−0.04***	−0.03**	−0.03**	−0.04**	−0.05**
	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
Peer network	0.04***	0.06***	0.10***	0.09***	0.10***	0.10***	0.03	0.06***	0.05*	0.07**	0.06*	0.08**
	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)	(0.04)	(0.04)
Indirect network effects	0.03***	0.04***	0.08***	0.07***	0.08***	0.08***	0.02***	0.04***	0.03***	0.05***	0.04***	0.06***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Residuals	0.04**	0.03					0.10***	0.09***				
	(0.02)	(0.02)					(0.02)	(0.02)				
Age	0.03***	0.03***	0.03***	0.03***	0.03***	0.03***	0.02***	0.02***	0.03***	0.02***	0.03***	0.02***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Age squared	−0.00***	−0.00***	−0.00***	−0.00***	−0.00***	−0.00***	−0.00***	−0.00***	−0.00***	−0.00***	−0.00***	−0.00***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Parent smoke	0.02**	0.01	0.02**	0.02**	0.02**	0.02*	0.04***	0.03***	0.05***	0.05***	0.05***	0.05***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Individual is male	0.13***	0.13***	0.14***	0.14***	0.14***	0.14***	0.16***	0.14***	0.18***	0.18***	0.18***	0.17***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Individual drinks alcohol often	0.13***	0.12***	0.29***	0.29***	0.29***	0.28***	0.17***	0.14***	0.35***	0.36***	0.34***	0.35***
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
Constant			−0.25***	−0.39***	−0.25**	−0.24**			−0.54***	−0.51***	−0.42***	−0.38***
			(0.10)	(0.07)	(0.10)	(0.10)			(0.08)	(0.08)	(0.10)	(0.10)
Observations	4,990	4,864	4,990	4,871	4,990	4,871	3,698	3,574	3,698	3,575	3,698	3,575
R-squared			0.50	0.48	0.50	0.50			0.38	0.41	0.39	0.43
Those with English as home language excluded	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes

Notes: The instrument used is the percentage of peer whose parents smoke. Standard errors are given in parentheses. Other controls include, race dummy, education dummy, religious dummy and parent education. Results from wave 3 are presented in columns (1), (2), (3), (4), (5) and (6), and wave 1 in column (7) to column (12). In columns (1), (3), (5), (7), (9) and (11), I exclude all individuals whose home language is English. Results of columns (1), (2), (7) and (8) are obtained from the control function approach, those in columns (3), (4), (9) and (10) are from the 2SLS approach without controlling for unobservables, and those in columns (5), (6), (11) and (12) are from the 2SLS controlling for unobservables. The dependent variable for all specifications is a dummy equal to 1 if the respondent's smokes. CF: control function approach. *Significant at the 10% level; **significant at the 5% level; and ***significant at the 1% level.

Table 4. Regression estimates of peer network as fixed effects are included

Variables	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Peer network on individual smoking intensity (Wave 3)						
Quantity of contacts	−0.23***	−0.17***	−0.13**	−0.22***	−0.15***	−0.12**
	(0.04)	(0.05)	(0.06)	(0.04)	(0.05)	(0.06)
Peer network	0.22***	0.17***	0.15***	0.20***	0.15***	0.14**
	(0.04)	(0.05)	(0.06)	(0.04)	(0.05)	(0.06)
Indirect network effects	0.19***	0.14***	0.12***	0.17***	0.12***	0.11***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Smoking addiction				0.03***	0.03***	0.03***
				(0.00)	(0.00)	(0.00)
Constant	0.35	0.42	0.12	0.73**	0.81**	0.43
	(0.32)	(0.35)	(0.42)	(0.33)	(0.35)	(0.42)
Observations	1,926	1,926	1,926	1,844	1,844	1,844
R-squared	0.10	0.12	0.12	0.13	0.15	0.16
District fixed effects	No	Yes	Yes	No	Yes	Yes
Language fixed effects	No	No	Yes	No	No	Yes
Panel B: Peer network on individual smoking intensity (Wave 1)						
Quantity of contacts	−0.08***	−0.06**	−0.07**	−0.08***	−0.07**	−0.08***
	(0.02)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)
Peer network	0.03***	0.02*	0.03*	0.03***	0.03**	0.03**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Indirect network effects	0.02***	0.01***	0.02***	0.02***	0.02***	0.02***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Smoking addiction				0.02***	0.02***	0.02***
				(0.00)	(0.00)	(0.00)
Constant	0.37*	0.39*	0.44	0.60***	0.66***	0.71**
	(0.20)	(0.23)	(0.29)	(0.21)	(0.23)	(0.29)
Observations	2,400	2,400	2,400	2,285	2,285	2,285
R-squared	0.11	0.15	0.15	0.12	0.16	0.17
District fixed effects	No	Yes	Yes	No	Yes	Yes
Language fixed effects	No	No	Yes	No	No	Yes

Notes: Standard errors are given in parentheses. Results from the 2SLS estimation are presented in columns (1)-(6). The results in Panel A obtained from individuals in wave 3 and the results in Panel B are from all individuals in wave 1. In columns (1) and (4), all possible fixed effects are excluded. In columns (2) and (5), district fixed effects are included. In columns (3) and (6), both district and language fixed effects are included. The dependent variable is the logarithm of the average number of cigarettes an individual smoke per day. Control variables include a Control variables including quadratic of age, dummies for race, dummies for education, dummies for religious, gender, drinking, parental education and parental smoking habit. *Denotes statistical significant at 10%, **denotes significant at the 5% level and ***denotes significant at the 1% level.

sensitivity check for the estimates of the preferred model specification (two-step control function approach). This preferred approach as indicated earlier provide a test for endogeneity of the peer network and more appropriate in nonlinear IV models. From the first column of Table 3, we can see that the reduced form residuals from the first step are significant in the structural equation. This indicates the existence of endogeneity of peer network. Therefore, results of the two-step control function approach, the 2SLS estimates or the fixed effects estimates are more plausible to the OLS estimates. Peer network remain positive and significant after the attempt to purge all possible biases.

While it is interesting that the estimated peer effects are statistically significant in both approaches, the magnitude of the peer effects from the 2SLS are generally larger than those of the control function approach. In addition, the probability of male smoking increases from 13% when the control function approach is used to 14% when the 2SLS is used (see column 1 and 3 for comparison). While drinkers are 13% more likely to smoke than non-drinkers in the CF approach, the propensity increases to 11% when the 2SLS is used (see column 1 and 3 for comparison). I

Table 5. Marginal effects and regression estimates of peer network as fixed effects are included

Variables	Wave 3				Wave 1			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Marginal effects estimates of peer network as fixed effects are included								
Quantity of contacts	−0.05***	−0.09***	−0.12***	−0.14***	−0.04***	−0.05***	−0.04**	−0.05***
	(0.01)	(0.01)	(0.02)	(0.03)	(0.01)	(0.02)	(0.02)	(0.01)
Peer network	0.04***	0.04***	0.04***	0.06***	0.10***	0.10***	0.10***	0.13***
	(0.01)	(0.01)	(0.01)	(0.02)	(0.03)	(0.04)	(0.04)	(0.02)
Indirect network effects	0.03***	0.03***	0.03***	0.04***	0.07***	0.07***	0.07***	0.10***
	(¡0.01)	(¡0.01)	(¡0.01)	(¡0.01)	(¡0.01)	(¡0.01)	(¡0.01)	(¡0.01)
Observations	5,533	5,533	5,533	2,931	4,579	4,579	4,579	2,385
District fixed effects	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Language fixed effects	No	No	Yes	Yes	No	No	Yes	Yes
Panel B: Regression estimates of peer network as fixed effects are included (aged 15 - 45)								
Quantity of contacts	−0.07***	−0.13***	−0.15***	−0.16***	−0.05***	−0.08***	−0.09***	−0.17***
	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)
Peer network	0.10***	0.12***	0.13***	0.20***	0.08***	0.09***	0.10***	0.34***
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Indirect network effect	0.08***	0.09***	0.10***	0.17***	0.06***	0.06***	0.07***	0.32***
	(¡0.01)	(¡0.01)	(¡0.01)	(¡0.01)	(¡0.01)	(¡0.01)	(¡0.01)	(¡0.01)
Constant	−0.24***	−0.27***	−0.19*	−0.65***	−0.25**	−0.18	0.03	−0.01
	(0.07)	(0.09)	(0.10)	(0.15)	(0.11)	(0.11)	(0.14)	(0.21)
Observations	10,883	10,883	10,883	3,698	11,223	11,223	11,223	4,990
R-squared	0.29	0.31	0.31	0.49	0.36	0.39	0.39	0.47
District fixed effects	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Language group fixed effects	No	No	Yes	Yes	No	No	Yes	Yes

Notes: Standard errors are given in parentheses. Results from wave 3 are presented in columns (1), (2), (3) and (4), and wave 1 in column (5) to column (8). The results in Panel A are marginal effects obtained from a logit model and for all individuals who are between the ages 15 and 24, and the results in Panel B are from a Linear Probability Model (LPM) for all individuals between the ages 15 and 45. In columns (1) and (5), all possible fixed effects are excluded. In columns (2) and (6), district fixed effects are included. In columns (3) and (7), both district and language fixed effects are included. In columns (4) and (8), all fixed effects and some parental characteristics are included. The dependent variable for all specifications is a dummy equal to 1 if the respondent has ever smoked. Control variables include a quadratic of age, dummies for race, dummies for education, dummies for religious, gender, drinking, parental education and parental smoking habit. ***Statistically significant at the 1% level; **statistically significant at the 5% level; *statistically significant at the 10% level.

further exclude individuals who reported English as their home language (though English is a home language to some individuals, it is also a medium of interaction common to almost everyone). The results for the remaining sample are presented in column 2 and column 4, and are not significantly different from those of the full sample. I also exclude the Afrikkans speakers since the share of smokers in the Afrikaans population is almost double their share in the general population and the results remain positive and statistically significant (see Table 8).

As an alternative peer network measure, wave 1 permits a further disaggregation of individuals in a given district into clusters. This measure can be considered more credible as it is more likely to assign individuals to their actual contacts than the district level measure. Table 6 reports the results obtain from this new measure of peer network. As expected, peer effects are generally higher when network size is measured at cluster than at district level. The peer network estimates for both measures are positive and statistically significant (see Table 2, Panel A of Table 5 and Panel A of Table 6 for comparison). The language-grouping effects are likely to vary across districts or clusters, and the

Table 6. Marginal and regression estimates of peer network after controlling for endogeneity

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Marginal and regression estimates of peer network (individuals aged 15–24)								
Quantity of contacts	−0.08***	−0.12***	−0.07***	−0.10***	−0.08***	−0.09***	−0.05***	−0.07***
	(0.01)	(0.01)	(0.01)	(0.03)	(0.01)	(0.01)	(0.01)	(0.02)
Peer network	0.07***	0.12***	0.13***	0.15***	0.07***	0.10***	0.10***	0.04***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)
Indirect network effects	0.06***	0.11***	0.12***	0.11***	0.06***	0.09***	0.09***	0.03***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Constant	−0.75**	−0.88***	−1.18***	−1.17***				
	(0.31)	(0.27)	(0.28)	(0.39)				
Residuals								0.02
								(0.02)
Observations	4,579	4,579	4,579	4,579	4,579	3,344	3,344	2.116
Cluster fixed effects	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Language fixed effects	No	No	Yes	Yes	No	No	Yes	Yes
District fixed effects	No	No	No	Yes	No	No	No	Yes
Interacted district and language fixed effects	No	No	No	Yes	No	No	No	Yes
R-squared	0.41	0.58	0.60	0.65				
Panel B: Marginal and regression estimates of peer network (individuals aged 15–45)								
Quantity of contacts)	−0.09***	−0.17***	−0.14***	−0.20***	−0.09***	−0.12***	−0.10***	−0.11***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)
Peer network	0.14***	0.20***	0.21***	0.25***	0.11***	0.14***	0.15***	0.25***
	(0.01)	(0.01)	(0.01)	(0.02)	(0.00)	(0.01)	(0.01)	(0.06)
Indirect network effects	0.13***	0.19***	0.21***	0.26***	0.10***	0.13***	0.14***	0.26***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Constant	−0.14***	−0.64***	−0.82***	−1.09***				
	(0.05)	(0.09)	(0.12)	(0.21)				
Residuals								0.12***
								(0.06)
Observations	10,883	10,883	10,883	10,883	10,883	10,565	10,565	3,445
Cluster fixed effects	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Language fixed effects	No	No	Yes	Yes	No	No	Yes	Yes
District fixed effects	No	No	No	Yes	No	No	No	Yes
Interacted district and language fixed effects	No	No	No	Yes	No	No	No	Yes
R-squared	0.44	0.55	0.56	0.63				

Notes: Standard errors are given in parentheses. Results from LMP are presented in columns (1), (2), (3) and (4), from the logit model in column (5) to column (7) and from the control function approach (IV results) in column (8). The results in Panel A are marginal effects obtained from individuals between the ages 15 and 24, and the results in Panel B are from all individuals between the ages 15 and 45. In this table, I present results from an alternative measure of peer networks where individuals are classified according to their respective clusters rather than districts. All results are obtained from wave 1 data set. In columns (1) and (5), all possible fixed effects are excluded. In columns (2) and (6), cluster fixed effects are included. In columns (3) and (7), both cluster and language fixed effects are included. In columns (4) and (8) cluster, language, district and interacted district and language fixed effects are included. The dependent variable for all specifications is a dummy equal to 1 if the respondent has ever smoked. Control variables include a control variables including quadratic of age, dummies for race, dummies for education, dummies for religious, gender, drinking, parental education and parental smoking habit. ***Statistically significant at the 1% level; **statistically significant at the 5% level; *statistically significant at the 10% level.

instrument may not be fully excludable from the second stage if there is a cultural effect embedded within it. I therefore include an interaction for the district and language fixed effects to control for the likely bias (see column 4 and 8 of Table 6 and column 4 to 6 of Table 8 for comparison).

Table 4 presents the peer network effects on an individual's daily smoking intensity for both waves. Similarly, the results suggest a significant peer effects on smoking intensity ranging from 14 to 22% in wave 3 and 2 to 3% in wave 1. Unlike the decision to

Table 7. Indirect network impact on smoking participation and smoking intensity

Variables	(1)	(2)	(3)	(4)	(5)
Panel A: Indirect network effects from Wave 1					
IsiNdebele	0.029	0.008	0.009	0.010	0.004
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
IsiXhosa	0.173	0.041	0.048	0.055	0.020
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
IsiZulu	0.227	0.052	0.062	0.071	0.025
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Sepedi	0.180	0.042	0.050	0.057	0.021
	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)
Sesotho	0.147	0.035	0.041	0.047	0.017
	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)
Setswana	0.183	0.043	0.051	0.058	0.021
	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)
SiSwati	0.153	0.037	0.043	0.050	0.018
	(0.005)	(0.001)	(0.001)	(0.002)	(0.001)
Tshivenda	0.178	0.042	0.049	0.057	0.020
	(0.007)	(0.002)	(0.002)	(0.002)	(0.001)
Xitsonga	0.067	0.017	0.020	0.023	0.008
	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)
Afrikaans	0.183	0.043	0.050	0.058	0.021
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
English	0.036	0.010	0.011	0.013	0.005
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Panel B: Indirect network effects from Wave 3					
IsiNdebele	0.014	0.005	0.010	0.012	0.016
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
IsiXhosa	0.085	0.027	0.062	0.070	0.102
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
IsiZulu	0.114	0.035	0.083	0.093	0.136
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Sepedi	0.090	0.028	0.066	0.074	0.107
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Sesotho	0.078	0.024	0.057	0.064	0.093
	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
Setswana	0.090	0.028	0.066	0.074	0.107
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
SiSwati	0.069	0.022	0.051	0.056	0.081
	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
Tshivenda	0.104	0.032	0.076	0.085	0.124
	(0.003)	(0.001)	(0.002)	(0.003)	(0.004)
IsiTsonga	0.027	0.009	0.020	0.022	0.031
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Afrikaans	0.093	0.029	0.068	0.076	0.110
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
English	0.010	0.003	0.007	0.008	0.011
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)

Notes: Standard errors are given in parentheses. Results from wave 1 are presented in panel A and Wave 3 in panel B. The results in column (1) are obtained from the peer network estimates in columns (4) and (8) of Table 2. The results in column (2) are obtained from peer network estimates in columns (2) and (8), those in column (3) from column (4) and (10) and those in column (4) from columns (6) and (12) of Table 3. The results in column (5) are obtained from column (6) of Table 4.

smoke, peer effects on smoking intensity are higher in wave 3 than in wave 1. However, it should be noted that estimates from both waves cannot be directly compared. The attrition level between wave 1 and wave 3, the unequal sample sizes, the changes in price and other unobservables could partly explain the huge differences in the magnitude of the estimates across waves. In addition to other controls in the decision to smoke estimation, I include smoking addiction (measured by the number of years an individual has been smoking cigarettes). The results suggest that smoking intensity increases with smoking addiction and the inclusion of addiction reduces the magnitude of peer networks.

Table 8. Regression estimates of peer network after excluding the Afrikaans sample

Variables	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Regression estimates for individuals between 15 and 45 years						
Quantity of contacts	−0.07***	−0.17***	−0.15***	−0.20***	−0.20***	−0.12***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.05)	(0.04)
Peer network	0.13***	0.22***	0.23***	0.28***	0.34***	0.33***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.09)
Residuals						−0.22**
						(0.09)
Constant	−0.28***	−0.22	−0.28	−0.69***	−1.09***	
	(0.03)	(0.18)	(0.19)	(0.26)	(0.24)	
Observations	9,101	9,101	9,101	9,101	3,698	2,910
Cluster fixed effects	No	Yes	Yes	Yes	Yes	Yes
Language fixed effects	No	No	Yes	Yes	Yes	Yes
District fixed effects	No	No	No	Yes	Yes	Yes
Interacted district & language fixed effects	No	No	No	Yes	Yes	Yes
R-squared	0.40	0.52	0.53	0.60	0.71	
Panel B: Regression estimates for individuals between 15 and 24 years						
Quantity of contacts	−0.04***	−0.14***	−0.10***	−0.14***	−0.14***	−0.05*
	(0.01)	(0.01)	(0.01)	(0.03)	(0.05)	(0.03)
Peer network	0.06***	0.13***	0.14***	0.16***	0.15***	0.02
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)
Residuals						0.03
						(0.03)
Constant	−0.60**	−0.65**	−0.81**	−2.28***	−1.76***	
	(0.27)	(0.32)	(0.32)	(0.44)	(0.45)	
Observations	2,075	2,075	2,075	2,075	2,417	1,801
Cluster fixed effects	No	Yes	Yes	Yes	Yes	Yes
Language fixed effects	No	No	Yes	Yes	Yes	Yes
District fixed effects	No	No	No	Yes	Yes	Yes
Interacted district & language fixed effects	No	No	No	Yes	Yes	Yes
R-squared	0.31	0.56	0.57	0.63	0.69	

Notes: Standard errors are given in parentheses. Results from LMP are presented in columns (1)-(5), and from the control function approach in column (6). The results in Panel A are estimates obtained from individuals between the ages 15 and 45, and the results in Panel B are from all individuals between the ages 15 and 24. All results are obtained from wave 1 data set, and clusters rather than districts are used to construct the peer network variable. In column (1), all possible fixed effects are excluded. In column (2), cluster fixed effects are included. In column (3), both cluster and language fixed effects are included, and in columns (4)-(6) cluster, language, district and interacted district and language fixed effects are included. In columns (1)-(4) parental characteristics are excluded. The dependent variable for all specifications is a dummy equal to 1 if the respondent has ever smoked. Control variables include a control variables including quadratic of age, dummies for race, dummies for education, dummies for religious, gender, drinking, parental education and parental smoking habit. ***Statistically significant at the 1% level; **statistically significant at the 5% level; *statistically significant at the 10% level.

6.1 Interpretation of the Network Estimates

The interpretation of the actual magnitude of the network estimates from the empirical analysis is not straightforward, given the way the network variable is computed. According to Bertrand *et al.* (2000), it is possible to arrive at a measure of the magnitude of the network effects, by asking to what extent social interaction would broaden a policy shock affecting the probability of smoking participation.[9] The argument is based on the

[9] I adopt the experimental approach, as specified in Bertrand *et al.* (2000) to identify the actual magnitude of peer network on smoking behaviour. First, it assumes a policy η which linearly affect smoking behaviour. The policy variable is included in the estimation, with the assumption that in the absence of peer effect, this variable is scaled such that a one percentage point rise in η will

Table 9. The first stage results: The effects of peer parents' smoking and education on peer smoking behaviour

Variables	Wave 1			Wave 3		
	(1)	(2)	(3)	(4)	(5)	(6)
Percentage of peer whose parent smoke	1.62***	1.40***	0.93***	2.21***	1.57***	0.37*
	(0.12)	(0.13)	(0.18)	(0.15)	(0.16)	(0.23)
Percentage of peer whose parent have tertiary education	−0.62***	−0.09	−0.02	0.07*	0.07*	0.05
	(0.20)	(0.20)	(0.25)	(0.04)	(0.04)	(0.04)
At least one parent smokes	0.09***	0.10***	0.10***	−0.04	−0.03	−0.04
	(0.03)	(0.03)	(0.03)	(0.06)	(0.06)	(0.06)
At least one parent has tertiary education	−0.11*	−0.06	−0.04	−1.85***	−2.48***	−3.03***
	(0.06)	(0.06)	(0.06)	(0.26)	(0.31)	(0.34)
Observations	3,698	3,698	3,698	4,990	4,990	4,990
Language fixed effects	No	Yes	Yes	No	Yes	Yes
District fixed effects	No	No	Yes	No	No	Yes
R-squared	0.49	0.51	0.58	0.43	0.46	0.53
F-statistics	207.62	142.51	64.42	235.41	159.78	71.01
Prob of *F*-stat	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Notes: Robust standard errors are in brackets. Control variables including quadratic of age, race dummy, education dummy, religious dummy, gender, drinking habit and quantity of contacts. The first stage estimates in columns (1)-(3) are for wave 1 sample and columns (4)-(6) estimates are for the wave 3 sample. The first stage estimates for peer parent smoking behaviour are similar across waves. The instrument is the percentage of peer whose parents smoke. *Denotes statistical significant at 10%, **denotes significant at the 5% level and ***denotes significant at the 1% level.

assumption that the policy shock is linear and the remaining marginal change after the policy shock is removed from the equilibrium outcome is attributed solely to peer effect $\left(\frac{1}{\left(1-\beta_1\overline{V_k}\right)}-1\right)$. The indirect network impact on the probability of smoking participation as well as smoking intensity are presented in all specifications as indirect network effects. The estimates of the indirect network effects presented in each table indicates the marginal change in smoking participation or smoking intensity induced solely by peer network. The network effects on smoking behaviour by language group are presented in Table 7. Peer network effects are generally and consistently higher among those who speak IsiZulu and lower among those speaking IsiNdebele than any other language group.

results to a one percentage point increase the probability of smoking, The estimation equation becomes: $Y_{ijt}=\eta+\beta_1(V_{jkt}*\overline{Y_{kt}})+\beta_2X_{ijt}+\beta_3P_{ijt}+\beta_4L_{jt}+\beta_5G_{jt}+V_{jkt}+\varepsilon_{ijt}$. Inclusion of the network variable generates a multiplier effect, such that in equilibrium, the increase smoking probability owing to the rise in η is higher. To illustrate this, I take the mean on both sides of the equation for each language group and differentiate with respect to η. In so doing I have, $\frac{d\overline{Y_{kt}}}{d\eta}=1+\overline{V_k*\frac{d\overline{Y_{kt}}}{d\eta}\beta_1}$ where $\overline{V_k}$ is the average number of contact $\left(V_{jk}\right)$ in each language group. The responsiveness of each language group's smoking probability, owing to the policy change, can be obtained by solving the derivation above for $\frac{d\overline{Y_{kt}}}{d\eta}$. In order to obtain the marginal change, resulting purely from peer network, I net out the direct effects of the policy change (note that it is equal to one). Hence, the actual magnitude of peer networks is given by $\frac{1}{\left(1-\beta_1\overline{V_k}\right)}-1$. Where $\frac{1}{\left(1-\beta_1\overline{V_k}\right)}-1$ is used to compute the indirect peer effect for each language group, and β_1 represent the respective network estimated coefficients.

Table 10. Regression estimates of peer network using education for network quantity (Wave 3)

Variables	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Quantity of contacts	−0.34***	−0.37***	−0.35***	−0.34***	−0.29***	−0.31***	−0.31***
	(0.02)	(0.02)	(0.02)	(0.03)	(0.04)	(0.05)	(0.05)
Peer network	0.79***	0.85***	0.85***	0.86***	0.49***	0.68***	0.68***
	(0.05)	(0.05)	(0.05)	(0.10)	(0.10)	(0.12)	(0.12)
Indirect network effects	0.44***	0.49***	0.49***	0.50***	0.23***	0.35***	0.35***
	(0.002)	(0.002)	(0.002)	(0.002)	0.001	(0.001	(0.001
Residuals					−0.02		
					(0.10)		
Constant	−0.34***	−0.28***	−0.39***	−0.41*		−0.46***	−0.46***
	(0.10)	(0.10)	(0.11)	(0.24)		(0.09)	(0.09)
Observations	11,310	11,310	11,310	5,002	5,002	5,002	5,002
R-squared	0.48	0.50	0.51	0.49		0.51	0.51
District fixed effects	No	Yes	Yes	Yes	No	No	Yes
Education fixed effects	No	No	Yes	Yes	No	No	Yes

Notes: Standard errors are given in parentheses. Non-IV results are presented in columns (1)-(4), IV results from the control function approach in column (5) and 2SLS results are presented in column (6) and (7). In columns (1), (5) and (6), all possible fixed effects are excluded. In column (2), district fixed effects are included. In columns (3), (4) and (7), both district and education fixed effects are included. In columns (1)-(3) parental characteristics are excluded. The dependent variable for all specifications is a dummy equal to 1 if the respondent has ever smoked. Control variables include a quadratic of age, dummies for race, dummies for education, dummies for religious, gender, drinking, parental education and parental smoking habit. ***Statistically significant at the 1% level; **statistically significant at the 5% level; *statistically significant at the 10% level.

7. CONCLUSIONS

In this paper, I deepen the empirical analysis of peer effects on cigarette smoking as presented in the literature by considering simultaneously their effects on the decision to smoke and on the smoking intensity in South Africa. Because cultural differences between countries may determine the extent to which smoking decision is influenced by peer (Gibbons *et al.*, 1995), and since the extent to which peer effect estimates for one country can be generalised to other countries has not been established (McVicar, 2011), I therefore provided evidence of peer effects on smoking propensities in South Africa. Specifically, I used a control function approach, a two-stage least square and/or a fixed effects approach to purge the potential biases from the endogenous peer effect estimates. This allows me to account for the problems of contextual effects, correlated effects and, simultaneity, identify the extent to which peer effect estimates rely on methodological approaches. Generally, the results indicate that peer network effects are quite robust to a series of alternative estimation approaches, across different dimensions of peer networks and different measures of smoking attitude.

Peer network estimates are positive and significant across all approaches, but the magnitude are found to vary with the estimation approach (see Tables 2 and 3). In addition, the estimates of peer effects are larger when age group 15–45 is considered than when 15–24 is used. Since the same data set and samples are used across the different methodologies, the differences in the magnitude of estimates may be readily interpreted as cross-method. It could also be as a result of age group variation in the magnitude of peer effects than is the case for differences in estimates across countries using different data as pointed out in McVicar (2011). While the variation in the magnitude of peer effects across methods could help explain the different challenges faced by each approach, the

variation across age groups could help explain the likely difference in social learning across age groups. The positive effect of peer networks suggests that policy interventions may have both direct and indirect (social multiplier) impact on an individual's smoking decision. Evidence of the indirect effects of price and other legislative policies are presented in Powell *et al.* (2005).

Evidence from this paper supports the theoretical underpinnings of peer influence discussed in Section 3. First the results show a direct influence of parents (see Table 9) and peers smoking attitude on individual smoking participation. This aligns with the social learning theory which states that behaviours are learned by observing others involved in a similar behaviour with parents and peers being the primary social factors. This also supports the primary socialisation theory which underscores that rational bonds between individuals and their family as well as friends are important in transmitting information about norms and behaviour. Second, individual characteristics such as education, religion, race and age among others are significant. This aligns with the social identity theory that focuses on an individual's self-concept and states that individual characteristics as well as those of the group matters. Precisely, the smoking habits of group members are likely to be similar if smoking is central to the social identity of the group. Finally, this paper uses the social network theory to identify one's network and confirm the interdependence of individuals and the existing rationale between them and their social system or neighbourhood.

Relative to the results of this paper, previous literature has documented larger peer effects on the decision to smoke. The fundamental question is: why are the peer effects on cigarette smoking low in South Africa relative other countries? A quick response may be; it is due to differences in the age group considered in this paper relative to that considered in most studies. This, to some extent, may reflect the biases of different strengths of networks in different countries. In addition and building on Gibbons *et al.* (1995), we might expect some correlation between country level cultural indicators and peer effect estimates that has been ignored in most studies but addressed in this paper.

It is acknowledged that the effectiveness of peer networks is contingent on differences in the characteristics of smokers, the characteristics of their contacts and/or their relationship with their contacts. It should be noted that in this paper, I do not ascertain the various channels through which the effectiveness of networks is contingent on. It simply illustrates the actuality of the network effects for respective social clusters. Justification of the effectiveness and pathways of operation of this peer effects require more detailed information on the relationship between smokers and their contacts. A further understanding of the dynamics and complexities of peer networks on smoking behaviour in South Africa hinge critically on a robust data set on peer networks, possibly on actual rather than potential contact. Although I am able to address the possible biases surrounding the estimation of peer networks, the nature of the data has limited the inclusion of an important age group (aged 10–14) that could be at a high risk of peer influence and birth order that might be correlated with peer networks. The data set has no information birth order among siblings limiting the inclusion of this variable in the estimation. The instrument is completely valid if the parents of the other geographical-language group does not influence one's own smoking behaviour. The data also limits the ability to ascertain this concern.

REFERENCES

ABEDIAN, I. and JACOBS, R. (2001). Tobacco taxes and government revenue in South Africa. *Journal of Economic Studies*, 28: 397-407.
ABRAMS, D. E. and HOGG, M. A. (1990). *Social Identity Theory: Constructive and Critical Advances*. New York: Harvester Wheatsheaf.
AJZEN, I. (1985). *From Intentions to Actions: A Theory of Planned Behavior. Action Control: From Cognition to Behaviour*, Heidelberg: Springer, 11-39.
AKERS, R. L. (1977). *Deviant Behavior: A Social Learning Approach* (2nd ed.). Belmont, CA: Wadsworth publishing.
ALBA, R. D. (1990). *Ethnic Identity: The Transformation of White America*. New Haven, CT: Yale University Press.
ALI, M. M. and DWYER, D. S. (2009). Estimating peer effects in adolescent smoking behavior: A longitudinal analysis. *Journal of Adolescent Health*, 45: 402-408.
ANGRIST, J. D. (2014). The perils of peer effects. *Labour Economics*, 30: 98-108.
ASARE, B. E. (2009). Tobacco regulation in South Africa: Interest groups and public policy. *African Journal of Political Science and International Relations*, 3: 99-106.
BALIA, S. and JONES, A. M. (2008). Mortality, lifestyle and socio-economic status. *Journal of Health Economics*, 27: 1-26.
BALTAGI, B. (2008). *Econometric Analysis of Panel Data* (Vol. 1). New York: Wiley.
BERTRAND, M., LUTTMER, E. F. and MULLAINATHAN, S. (2000). Network effects and welfare cultures. *The Quarterly Journal of Economics*, 115: 1019-1055.
BOSHOFF, W. H. (2008). Cigarette demand in South Africa over 1996–2006: The role of price, income and health awareness. *South African Journal of Economics*, 76: 118-131.
BURNS, J., GODLONTON, S. and KESWELL, M. (2010). Social networks, employment and worker discouragement: Evidence from South Africa. *Labour Economics*, 17: 336-344.
CHOU, S.-Y., LIU, E., LIN, M.-J. and LIU, J.-T. (2015). Better peers, better scores? A study of twin junior high school graduates in Taiwan. *Applied Economics*, 47: 3462-3481.
DAHL, G., LOKEN, K. and MOGSTAD, M. (2014). Peer effects in program participation. *American Economic Review*, 104: 2049-2074.
DERI, C. (2005). Social networks and health service utilization. *Journal of Health Economics*, 24: 1076-1107.
DUARTE, R., ESCARIO, J.-J. and MOLINA, J.-A. (2014). Are estimated peer effects on smoking robust? Evidence from adolescent students in Spain. *Empirical Economics*, 46: 1167-1179.
EISENBERG, D. (2004). *Peer Effects for Adolescent Substance Use: Do They Really Exist*. UC-Berkeley School of Public Health Working Paper.
FISHBEIN, M. and AJZEN, I. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley Publishing Co.
FLETCHER, J. (2010). Social interactions and smoking: evidence using multiple student cohorts, instrumental variables, and school fixed effects. *Health Economics*, 19: 466-484.
FLETCHER, J. M. (2012). Peer influences on adolescent alcohol consumption: Evidence using an instrumental variables/fixed effect approach. *Journal of Population Economics*, 25: 1265-1286.
FOWLER, J. H. and CHRISTAKIS, N. A. (2008). Estimating peer effects on health in social networks: A response to Cohen-Cole and Fletcher; and Trogdon, Nonnemaker, and Pais. *Journal of Health Economics*, 27: 1400-1405.
GIBBONS, F. X., HELWEG-LARSEN, M. and GERRARD, M. (1995). Prevalence estimates and adolescent risk behavior: Cross-cultural differences in social influence. *Journal of Applied Psychology*, 80: 107.
GOVERNMENT, S. (2007). Tobacco product control amendment act. *Government Gazette*, 512.
GRANOVETTER, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, 78: 1360-1380.
GROENEWALD, P., VOS, T., NORMAN, R., LAUBSCHER, R., Van WALBEEK, C., SALOOJEE, Y., SITAS, F., BRADSHAW, D. and South African Comparative Risk Assessment Collaborating Group. (2007). Estimating the burden of disease attributable to smoking in South Africa in 2000. *South African Medical Journal*, 97: 674-681.
HAWKINS, J. D. and WEIS, J. G. (1985). The social development model: An integrated approach to delinquency prevention. *Journal of Primary Prevention*, 6: 73-97.
HIRSCHI, T. (1969). *Cause of Delinquency*. Los Angele, CA: University of California Press.
HORRACE, W. C. and OAXACA, R. L. (2006). Results on the bias and inconsistency of ordinary least squares for the linear probability model. *Economics Letters*, 90: 321-327.
KOBUS, K. (2003). Peers and adolescent smoking. *Addiction*, 98: 37-55.
KRAUTH, B. V. (2007). Peer and selection effects on youth smoking in California. *Journal of Business & Economic Statistics*, 25: 288-298.
KWON, I. and JUN, D. (2015). Information disclosure and peer effects in the use of antibiotics. *Journal of Health Economics*, 42: 1-16.
LEAVER, R. (2002). Tobacco control policy in South Africa as seen through the eyes of the media. In C. van Walbeek (ed.), *The Economics of Tobacco Control in South Africa*. University of Cape Town.
MANSKI, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, 60: 531-542.
McVICAR, D. (2011). Estimates of peer effects in adolescent smoking across twenty six European countries. *Social Science & Medicine*, 73: 1186-1193.

——— and POLANSKI, A. (2014). Peer effects in UK adolescent substance use: Never mind the classmates?. *Oxford Bulletin of Economics and Statistics*, 76: 589-604.
MUKONG, A. K. and BURNS, J. (2015). *Social Networks and Maternal Health Care Utilisation in Tanzania*. Working Paper 540, Economic Research Southern Africa (ERSA).
NAKAJIMA, R. (2007). Measuring peer effects on youth smoking behaviour. *The Review of Economic Studies*, 74: 897-935.
NORTON, E. C., LINDROOTH, R. C. and ENNETT, S. T. (1998). Controlling for the endogeneity of peer substance use on adolescent alcohol and tobacco use. *Health Economics*, 7: 439-453.
OETTING, E. and BEAUVAIS, F. (1986). Peer cluster theory: Drugs and the adolescent. *Journal of Counseling and Development*, 65: 17-22.
OETTING, E. R. and DONNERMEYER, J. F. (1998). Primary socialization theory: The etiology of drug use and deviance. I. *Substance Use & Misuse*, 33: 995-1026.
OREOPOULOS, P. (2003). The long-run consequences of living in a poor neighborhood. *The Quarterly Journal of Economics*, 118: 1533-1575.
PARKER, J. (2012). Does living near classmates help introductory economics students get better grades? *The Journal of Economic Education*, 43: 149-164.
PAXTON, P. (1999). Regression models for categorical and limited dependent variables, sage advanced quantitative techniques in the social sciences series. *Social Forces*, 77: 1245-1246.
POWELL, L. M., TAURAS, J. A. and ROSS, H. (2005). The importance of peer effects, cigarette prices and tobacco control policies for youth smoking behavior. *Journal of Health Economics*, 24: 950-968.
RIVERS, D. and VUONG, Q. H. (1988). Limited information estimators and exogeneity tests for simultaneous probit models. *Journal of Econometrics*, 39: 347-366.
SILLES, M. (2015). The causal effect of schooling on smoking behavior. *Economics of Education Review*, 48: 102-116.
SIMPSON, J. H. (2000). Social learning and social structure: A general theory of crime and deviance. By Ronald L. Akers. Northeastern University Press, 1998. 420 pp. *Social Forces*, 78: 1171-1173.
SOETEVENT, A. R. and KOOREMAN, P. (2007). A discrete-choice model with social interactions: With an application to high school teen behavior. *Journal of Applied Econometrics*, 22: 599-624.
Van WALBEEK, C. (2000). Impact of the recent tobacco excise tax increases on the future government revenue potential in South Africa. *Economics of Tobacco Control Project.*
——— (2002). Recent trends in smoking prevalence in South Africa–some evidence from AMPS data. *South African Medical Journal*, 92: 468-472.
——— (2004). Tobacco control in South Africa in the 1990s. A mix of advocacy, academic research and policy. *South African Journal of Economic History*, 19: 100-131.
Van WALBEEK, C. P. (1996). Excise taxes on tobacco: How much scope does the government have? *South African Journal of Economics*, 64: 20-42.
VARDARDOTTIR, A. (2013). Peer effects and academic achievement: A regression discontinuity approach. *Economics of Education Review*, 36: 108-121.
WASSERMAN, S. and FAUST, K. (1994). *Social Network Analysis: Methods and Applications* (Vol. 8). Cambridge: Cambridge university press.
WORLD HEALTH ORGANIZATION (2015). WHO report on the global tobacco epidemic, 2015: Raising taxes on tobacco. Geneva, Switzerland: World Health Organization
WOOLDRIDGE, J. M. (2010). *Econometric analysis of cross section and panel data*. Cambridge: MIT Press.

APPENDIX

The 2SLS and the generalised method of moments (GMM) are the standard IV methods for estimating models linear in parameters. The control function (CF) approach which hinges on the same identification conditions like the 2SLS and the GMM is and alternative. Like the 2SLS, in the CF approach we first estimate the model of endogenous regressors as a function of instruments and use the errors as an additional regressor in the main model. First, consider a binary outcome model:

$$Y = M(X, \beta, \varepsilon) \tag{2}$$

Where Y is a binary response variable, X is a vector of explanatory variables including the endogenous variable(s), β is a vector of parameter estimates and ε is the error term. Let G, h and a well-behaved error μ be assumed function from equation (2) such that $X^{\rho} = G(Z, \rho)$, $\varepsilon = h(\rho, \mu)$ and $\mu \perp (X, \rho)$. Where X^{ρ} is the endogenous variable, Z is an exogenous variable correlated with X^{ρ} and not with Y and ρ are derived fitted values of the

error term. The first step estimate the endogenous regressor $G(Z, e)$, then at the second step we estimate:

$$Y = M(X, \beta, h(\rho, \mu)) = \tilde{M}(X, \rho, \beta, \mu) \tag{3}$$

Where μ is the error term to the $\tilde{M}$ model. The error term is independent of (X, ρ) and the model no longer has an endogeneity problem and can be estimated using the probit method. The CF approach require the endogenous regressor $X^{\rho}$to be continuous and not applicable for binary, discrete or censored endogenous regressors. For example, the assumption that the errors are normally distributed and independent of the regressors at the first stage will not be satisfied.