STOCKHOLM SCHOOL OF ECONOMICS
Department of Economics
659 Degree Project in Economics
Spring 2016

Extending the Grossman model: the association between societal apprehension and investments in health capital

Mamud Miyan (22237) Alexander Uggla (23058)

ABSTRACT: The Grossman model is a widely used theoretical framework in the field of health economics. However, it has been criticized on the basis of its assumptions being difficult to apply outside of theory. Consequently, several attempts have been made at extending and improving the model. In this thesis, we propose yet another extension in an effort to increase the applicability of the framework. The extension is motivated from empirical observations in the medical field suggesting that an increase in the degree of apprehension on an individual level increases healthcare-seeking. This essay investigates whether apprehension on a societal level may influence investments in health capital–a crucial part of the Grossman model. We use emergency department attendance as a proxy for investments in health capital, and two indices representing societal apprehension are created based on the quantity of apprehension-reflecting vocabulary used in news media and social media respectively. Controlling for previously known emergency department attendance determinants, we find that societal apprehension significantly influences the amount of daily patient visits. On average, we find that the daily healthcare-seeking behavior of 23 and 8 persons can be explained by an average daily change of societal apprehension, as measured by the two indices respectively. We conclude that societal apprehension significantly affects a person's tendency to invest in health capital through a change of the perceived cost of capital and/or a shift in the MEI-curve, and as such constitutes a reasonable extension of the Grossman model.

Keywords:	Grossman model, health capital, societal apprehension, news media, social media
JEL:	I1, I12
Internal supervisor:	Martina Björkman Nyqvist
External supervisor:	Martin Nordberg
Date submitted:	May 16, 2016
Date examined:	May 30, 2016
Discussants:	Julia Lundholm, Richard Sunnergårdh
Examiner:	Örjan Sjöberg

Acknowledgments

We would like to express our gratitude to our supervisor Martin Nordberg, medical doctor at Stockholm South General Hospital and PhD candidate at Karolinska Institutet, for his major commitment and expertise and to our supervisor Martina Björkman Nyqvist for her guidance. In addition, we wish to thank Per Landin and others at Retriever, as well as Gavagai, for providing us with data and insights. We would also like to thank Erik Lindqvist and Per-Olov Edlund for their assistance during statistical analysis.

Table of contents

1. Introduction

The Grossman model has arguably been the most prominent and influential model in the field of health economics since its creation in 1972 (1, 2). In this model, health is considered a type of durable capital stock that produces healthy time as an output. It is assumed that individuals initially possess a stock of health capital that can be increased by engaging in health-promoting activities (2).

Overall, the Grossman model has significantly deepened our understanding of how individuals distribute their resources to produce health and there is empirical evidence for several predictions made by the model (1). However, several authors have also criticized the model and questioned its underlying assumptions (3, 4, 5-7). Multiple improvements and extensions have been suggested to overcome its limitations (3, 4, 6-8). In this essay, we explore and evaluate yet another potential refinement and extension, namely the role of societal apprehension in determining investments in health capital. Empirical observations in the medical field has long recognized that a significant proportion of emergency ward patients are in fact not in need of emergency medical services (9). Consequently, the question arises: what constitutes the trigger that brings these patients to the emergency department (ED)? We use emergency department attendance as a proxy for investments in health capital, and investigate the association between the degree of societal apprehension and the number of emergency department patient visits. Two apprehension indices are created using the quantity of apprehension-reflecting vocabulary expressed in influential Swedish media outlets and in predominantly social media respectively. Subsequently, the indices are used as a measure of societal apprehension. Our research highlights a so far overlooked aspect affecting investments in health capital–a cornerstone of the Grossman model–and it therefore has the potential to increase the empirical accuracy and utility of the model. We hypothesize that an increased degree of societal apprehension is associated with an increased tendency to invest in health capital.

This essay is structured as follows. In the next section, we briefly summarize the Grossman model. Following this, we familiarize the reader with our study environment. Next, there is a section covering previous research regarding both the Grossman model with its critique and extensions as well as regarding emergency department attendance, societal apprehension and the media's effect on our behavior. Succeeding this, we outline our hypotheses, describe our data, and present the empirical methodology used. Section nine is devoted to a discussion of our findings. We end the paper in section ten by outlining the contributions of this study in a broader context, some concluding remarks, and suggestions for future research.

2. The Grossman model

In this section, we explain the fundamentals of the Grossman model and thereby present the theoretical framework underlying this essay on which we later develop our hypotheses.

In 1972, Michael Grossman presented a model for the production of health through the demand of health capital, which subsequently came to be named after him. The model is originally based on human capital theory, according to which people invest in themselves through education, training and health in order to increase future earnings (2).

There is an inherent demand for health due to several reasons: people feel better when healthy (less time is lost to illness, and hence more work and income can be earned), productivity increases with health, and finally, people may live longer. In order to achieve these benefits, investments in health capital are made. According to Grossman, these investments consist of health-improving behavior and measures such as exercising, maintaining a wholesome diet, and medical care, as illustrated in Figure 2 (1).

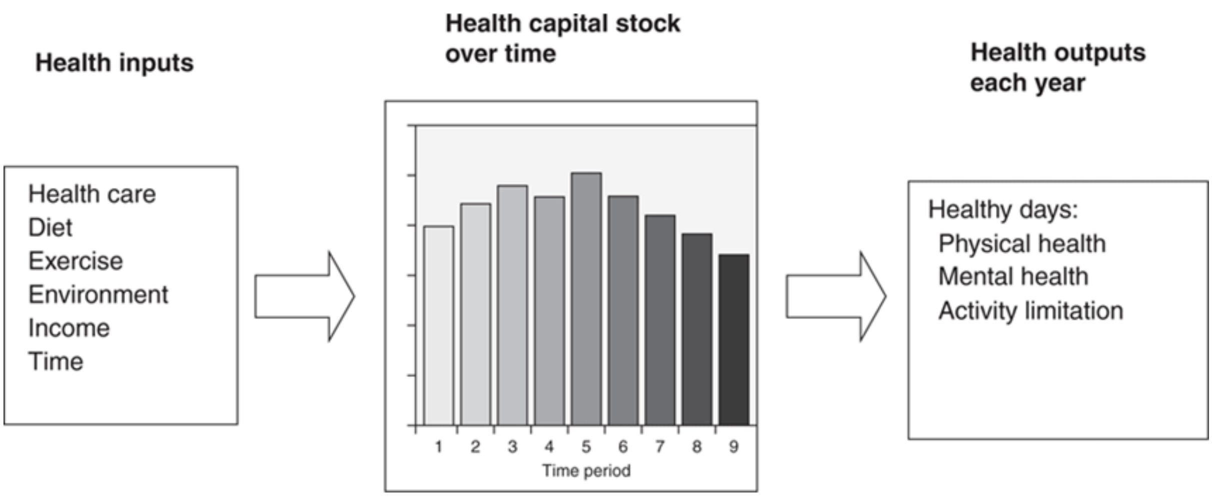

Figure 2: Inputs, health capital stock, and its outputs. Reproduced from Folland et al. (1).

However, the demand for health has some distinguishable features as compared to the demand of other goods:

- The consumer does not desire medical care in itself, but rather health–the outcome of medical inputs
- The consumer simultaneously acts as a consumer and producer by consuming as well as allocating time and resources to health-promoting activities
- Health lasts for more than one period and does not depreciate instantly
- Perhaps most importantly, health constitutes both a consumption good (it is desired since health capital increases well-being) and an investment good (it is yearned for because it increases the number of healthy days available to work and earn income)

2.1 The trade-off between health, time, and other goods

An individual has a finite amount of time. Thus, investing in health as a capital stock, albeit being desirable, involves trading off time and other resources. This gives rise to the production possibility frontier, as seen by A-E-C-D in Figure 2.1a. Depending on the relative importance of health as opposed to other goods, such as bread in Figure 2.1a, the indifference curves change. If health is solely an investment good and utility is derived from bread only, the indifference curve will be depicted by U_1. On the other hand, if health is a consumption good, and utility is derived from both health and bread, the indifference curve will be represented by U_2 (1).

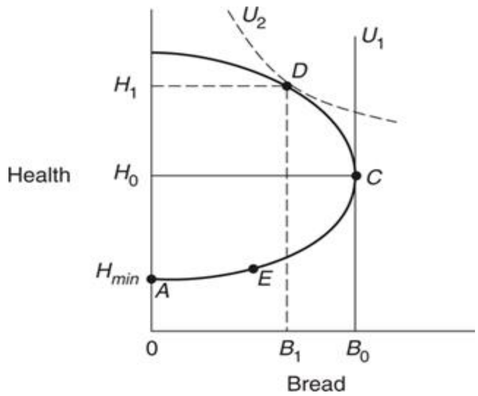

Figure 2.1a: Allocation of resources between bread and health. Reproduced from Folland et al. (1).

According to Grossman's reasoning, as an individual invests in health capital, less illness is experienced. Consequently, total available time for both leisure and work (which provides income) increases, as illustrated in Figure 2.1b. An outward shift of the income-leisure trade-off curve is observed from VS to RQ, entailing an increased utility E'. Hypothetically, more health stock may also correlate positively with productivity, resulting in a steeper curve (not seen in the figure) (1).

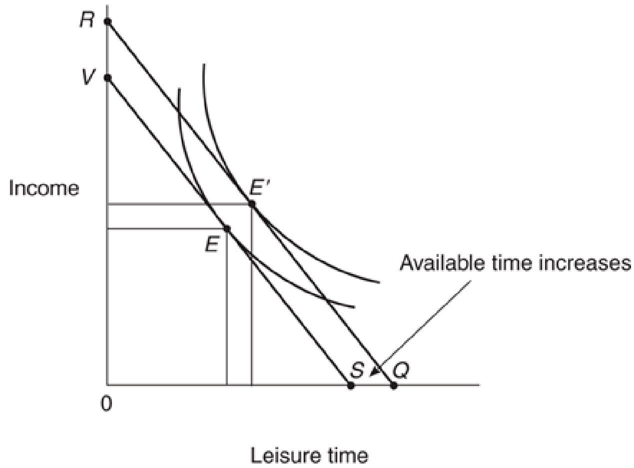

Figure 2.1b: Increased amount of healthy time due to investments in health capital stock. Reproduced from Folland (1).

However, increasing health investments do not always increase both leisure and income, as the production of healthy days is subject to the law of diminishing returns: additional resource and time investments in health capital stock have smaller impacts on the marginal output of healthy days. This gives rise to the concept of marginal efficiency of investment (MEI). The decreasing MEI is represented by the flattening of the curve in Figure 2.1c, as successively larger increases in health stock (on the horizontal axis) are required to achieve equivalent increases in output of healthy days (on the vertical axis) (1).

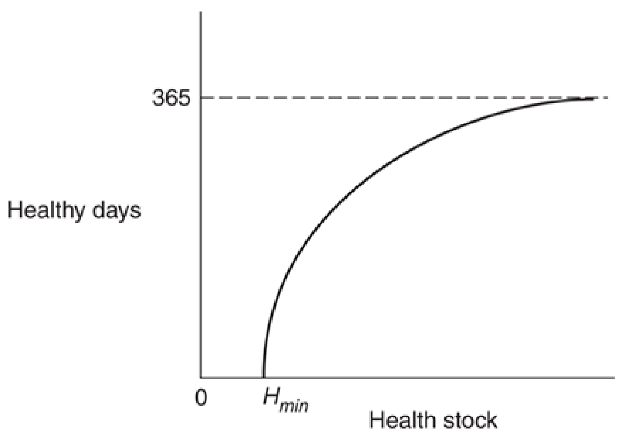

Figure 2.1c: Relationship of healthy days to health stock. Reproduced from Folland (1).

2.2 Optimal health stock

The optimal health capital stock for an individual depends on the cost of capital health, which denotes the combined effect of an opportunity cost (foregone resources in terms of money and time) and the depreciation rate (the pace at which health deteriorates) (1). Formally:

$$\textit{Cost of capital} = \textit{interest rate} + \textit{depreciation rate} = r + \delta$$

The optimal level of health stock is at the intersection of the MEI-curve and the cost of capital in Figure 2.2. The consumer of health chooses an optimal, equilibrium level of health stock by deciding how much resources to allocate to work, health-improving measures (such as diet, exercises, physician checkups, etc.) and hobbies. The total time available for the individual constitutes the major restraint. In the model, resources are allocated so that the optimal level of health stock is preserved each year, resulting in an equilibrium level of healthy days per year. At H_{min}, the production of healthy days is zero, indicating death. Therefore, it is necessary to raise H above H_{min} to obtain income and leisure time that may be devoted to hobbies or further health investments (1).

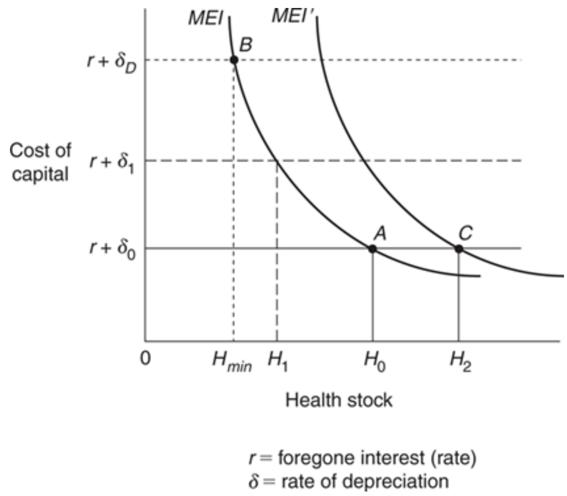

Figure 2.2: Optimal health stock with different cost of capital. Reproduced from Folland (1).

2.3 Changes in health stock equilibrium

Figure 2.2 also helps us understand causes of health stock equilibrium changes. Equilibrium changes are brought on by changes in several variables, including age, education, and wage. As we age, our health deteriorates faster, resulting in a higher cost of capital of health, and thus a falling optimal level of health stock (1). In Figure 2.2, this is illustrated as the intersection of the MEI-curve occurs at a H_1 instead of H_0 as the cost of capital increases from $r + \delta_0$ to $r + \delta_1$. Further, with a higher wage, the return on healthy days increases, causing a shift of the MEI-curve to the right (from MEI to MEI') and a higher optimal health stock. Similarly, a higher education improves the efficiency of producing health, resulting in an analogous shift to the right of the MEI-curve. Thus, several factors affect the equilibrium health capital stock and thus the tendency to invest in health (1, 3). In this essay, we hypothesize and test whether apprehension on a societal level may be yet another such factor.

3. Background

Below, we introduce the reader to our study environment.

Healthcare in Sweden is organized by county councils and all inhabitants are covered by the national health insurance, which is mainly financed by taxes (9). Patients pay a small amount for each healthcare visit: in Stockholm County Council, adult patients pay a fee of 400 SEK to visit an ED up to a ceiling of 1100 SEK yearly. The ceiling of 1100 SEK includes all healthcare expenditures, not only ED visits. Any cost above this is entirely covered by the county council (10).

The national healthcare system is further organized into inpatient care, emergency care and primary care. Primary care is principally intended for non-urgent issues and patients are referred to a specialist if the condition is assessed to be too severe or the resources inadequate for optimal patient care. In contrast, emergency departments are intended for urgent issues. Considering that the Swedish healthcare system is publically financed, it operates with limited resources and is assigned to allocate these optimally.

3.1 Emergency departments in Stockholm County Council

Emergency medical services are provided at seven hospitals in Stockholm County Council: Karolinska University Hospital (divided into two separate units in Solna and Huddinge respectively), Danderyds Hospital, Stockholm South General Hospital, Capio St. Görans Hospital, Södertalje hospital, and Norrtälje Hospital. In addition, our analysis includes data from St. Erik

Eye Hospital in central Stockholm and the county council-enclasping maxillofacial surgery emergency department. In total, these hospitals comprise a catchment area of more than two million inhabitants and several hundred thousand yearly patient visits, with a complete comprehension of all medical fields except psychiatry. The demand for psychiatric emergency care was deemed to be inherently different from somatic fields of medicine, and was consequently excluded from the analysis. At all hospitals except Soder General, patients are initially assessed using a triage system, where vital parameters and the patients' previous medical history are taken into consideration by a nurse to stratify patients according to level of urgency. Subsequently, patients are prioritized accordingly by physicians (11). In contrast, Soder General Hospital recently implemented new routines where patients are initially assessed by a physician before receiving a priority level.

4. Previous research

In this section, we compile and outline research concerning the Grossman model's limitations and suggested improvements. Also, we summarize previous research regarding determinants of emergency department attendance, and media's (on which our apprehension indices are based) effect on apprehension and health-related behavior. As such, we outline a theoretical as well as an empirical justification for our study.

4.1 Critique and extensions of the Grossman model

Although the Grossman model has largely appealed to economists, it has failed to achieve approval from both laypersons and policy makers. The original model has remained relatively unchanged since its introduction (12). In addition, empirical testing of the model has proven arduous due to the imperceptible nature of some variables (4). In this section, we present an overview of research highlighting extensions and adjustments that have been proposed in order to clarify controversial parts of the model.

According to Hren (2011), the Grossman model severely neglects the element of uncertainty. As Hren puts it, the model is *"deterministic and does not take into account the often observed random occurrences of illnesses"* (4). In reality, the presence of uncertainty is known to largely affect the health investment behavior of individuals (3, 6). Further, the model assumes that the demand for health is relatively constant, yet in reality demand spikes at times of sudden illness. Thus, health demand actually correlates negatively with current health status, which is often explained by people not worrying about one's health until actually falling ill (4, 5). Closely related to uncertainty, the model also neglects imperfect information. On a theoretical level, the model assumes that individuals possess complete and flawless information regarding their health capital, marginal benefits of investments, and changes in depreciation rate. As such, it again harshly simplifies the stochastic and unpredictable nature of disease. All in all, Hren states that its theoretical and empirical shortcomings may undermine the applicability of the model (4).

Further, Bolin et al. investigated interactions between spouses and family members regarding health inputs, and found possibilities of family members free riding on one another (8). Empirically, it is well known that an individual's health closely depends on one's social surroundings. Low socioeconomic status is a known risk factor for both somatic and psychiatric illness (13). Thus, Bolin illustrated a significant flaw of the model: looking at individuals' health in isolation, without considering the social environment. In proximity with the role of the family, Galama et al. proposed the incorporation of a childhood phase into the model (14). According to the authors, childhood health capital endowments significantly influence future health outcomes and as such, ought to be regarded in the model.

It is also worth noting that the Grossman model can be extended in an unchanged form into fields beyond its original use. For example, Michael in 2004 introduced the concept of sexual capital and successfully implemented the model's framework in a partially new field (7).

In conjunction, the extent and magnitude of the proposed extensions of the model illustrate its shortcomings and the potential for improvement. Consequently, we investigate yet another conceivable extension, one which may influence investments in health capital–societal apprehension.

4.2 Emergency department attendance

Considering the crucial role of the emergency department in a modern healthcare organization, a large line of research has been dedicated to studying determinants of its attendance. Below, we review the empirical evidence regarding ED attendace determinants, which serves as a justification for subsequent control-variables when statistically investigating the influence of apprehension on investments in health capital. As previously mentioned, investments in health capital are in this thesis proxied by emergency department visits.

4.2.1 Overcrowding issues in the emergency department

The emergency department constitutes an advanced medical setting where both life-threatening and routine acute medical conditions are managed continuously on a daily basis. Also, it functions as a gateway for hospital admittance for patients arriving by ambulance, by referral from primary care facilities, as well as for patients arriving by own initiative. Simultaneously, the emergency setting comprises a complex environment, involving expertise from multiple fields, where several professions interact.

The demand for emergency medical services has increased significantly in the last few years, resulting in overcrowding and increased healthcare costs at several emergency departments (15). For example, 47 % of American hospitals reportedly operated at a level exceeding or equivalent to their ED capacity in 2007 (16). Overcrowding is known to cause decreased quality of treatment and prognosis, and hence constitutes an impediment to optimal patient care (15). In order to address the overcrowding, multiple measures have been attempted with variable success including staffing adjustments, opening of light emergency clinics and expansion of the number of beds and spaces. It has previously been established that knowledge regarding what affects the timing and magnitude of overcrowding, as well as the ability to accurately anticipate these conditions, carry the potential to significantly improve ED operations (15, 17).

Despite overcrowding, it is well known that a significant proportion of emergency ward patients are not in emergent need of medical care. These are known as non-urgent patients. In fact, up to 55 % of emergency departments' resources may be occupied by non-urgent patients (9). Studies have shown that one partially responsible cause of non-urgent patients seeking emergency care is a state of apprehension amongst individuals (9, 18). To our knowledge however, the role of apprehension on a societal level in explaining ED attendance has not yet been studied.

4.2.2 Determinants of ED attendance in general

As a result of overcrowding issues, previous research has been aimed at predicting emergency department attendance and identifying its determinants. Several variables have been investigated, such as day of the week, day in relation to holiday, month of the year, weather, and daily temperature (15, 19-27). In addition, a group of Swedish authors recently investigated patients' health related web searches the day before visiting the ED and its relation to overall attendance (28).

Regarding calendar variables, empirical observations most commonly suggest a weekly pattern with the highest amount of visits on Mondays, and thereafter a declining trend with the fewest patients in the ED on Sundays (19-21). A recent systematic review found that day of the week had the strongest effect of all the variables examined in explaining ED patient volume (26). In addition, two studies have found that the day after a holiday was significantly and independently associated with approximately an 11 % increase in patient volume (19, 21). As for monthly distributions, different studies report contradictory results. While Diehl et al. and Glass reported the highest ED attendance throughout the summer months, Batal as well as Holleman found that winter months were the busiest (19-21, 27).

The role of weather and temperature in determining ED attendance has also been discussed. Opposing results have been reported, and a systematic review determined that the addition of meteorological data had no additive value in models forecasting the number ED visits (15, 17, 19-21). Other temporal variables, including local or international events, may also affect the number of patient visits (26). For instance, Redelmeier et al. found a 17 % decrease of patient visits during an Olympic gold medal television broadcast, as compared to the same weekday three weeks before and after the broadcast (29).

Furthermore, at least one study has researched the relationship between online information seeking and ED attendance. Ekström et al. investigated how the number of visits to a regional medical web site, the Stockholm Health Care Guide, correlated with ED attendance the following day. Web site visits were viewed as a proxy for the health concern of the public. The study concluded that there was a strong and significant correlation between web site visits during 6 p.m. to midnight and ED attendance the next day (28).

4.3 The role of mass media

In order to develop and test our hypothesis, we must understand mass media's influence on health-seeking behavior. This is because, as mentioned earlier, we use media activity based indices as proxies for societal apprehension when studying its relationship to investments in health capital.

4.3.1 Media campaigns and health-seeking behavior

It has been well documented that mass media campaigns and media coverage can have an effect on health-seeking behavior (30-32). In a literature review, Noar retrospectively examined research regarding public health campaigns, conducted over a ten-year period. The author found evidence that mass media campaigns regarding health issues do have an effect on health-seeking behavior. However, Noar stresses that such campaigns must be targeted and well executed to have effect (31). Another example is provided by Flowers et al., who investigated the relationship between a mass media campaign and the frequency of which persons affected by the campaign attended a HIV testing clinic. It was concluded that those to a higher degree exposed to the campaign were more likely to attend a clinic than those exposed to a lesser degree (32). Furthermore, in a comprehensive review by Grilli et al., twenty publications on the topic were examined. Fifteen of these regarded mass media campaigns, while five investigated media coverage concerning a health issue in general. The authors conclude that although a few of the examined publications lacked statistical analysis or used statistical analysis in an inappropriate way, there is evidence that mass media does have a measureable impact on health service attendance (30).

4.3.2 Mass media and apprehension

As media can influence health-related behavior, it seems likely that it can also impact behavior and sentiment in issues not related to health. Indeed, it has been observed from a wide array of contexts that mass media are able to exert strong, long-term effects on audiences and shape perceptions. The term *agenda setting* has been established to describe the correlation between mass media's emphasis on particular issues and the importance credited to these issues by the public (33). In *Agendas, Alternatives, and Public Policies*, Kingdon argues on the basis of empirical case studies that mass media can affect public opinions and magnify the perceived importance of events (34). Further, mass media's presentation of an issue significantly influences how it is interpreted by audiences, a phenomenon known as *framing*, with theoretical roots in psychology as well as sociology (33). Kahneman received the 2002 Nobel Prize in economics for his studies on how different presentations of identical situations influence people's choices and evaluations of the available alternatives (35).

In line with mass media's proven ability to affect the sentiment of its audiences, further research has shown that mass media also has significant ramifications within the field of public health. News reporting may markedly influence people's perception of health issues and induce fear of disease (36, 37). Notably, following Nancy Reagan's choice to undergo a mastectomy (where the whole breast is removed) in 1987 and the subsequent media reporting, a 25 % decrease in women choosing to undergo breast-conserving surgery was noted in the US (38). More recently, the highly publicized Ebola cases have been argued to cause a disproportionate contagion of fear (39). In fact, an association between media reporting of an event and consequent suicidal behavior, also known as *imitative suicides*, have been described and named *The Werther effect* (40).

4.4 Apprehension and healthcare seeking

The association between apprehension and healthcare seeking behavior has mostly been studied indirectly. For example, Backman et al. studied the characteristics of non-urgent patients at emergency departments and concluded that these were more anxious and disturbed by their symptoms compared to patients who used non-scheduled appointments at primary care facilities (9). It thus seemed as if the most concerned patients sought what they perceived to be the most advanced healthcare institution available. Furthermore, parental concerns are known to be a cause of increased emergency department visits among pediatric patients (40). In addition, Filipkowski et al. reported that concerns over modern technology, such as air pollution, X-rays, and food additives, influenced the number of visits to healthcare providers even among young and healthy people (18). As apprehension on an individual level clearly affects health-seeking behavior and thus the individual's inclination to invest in health capital, apprehension on a societal level seems to be a reasonable extension of the Grossman model.

5. Hypotheses

The aim of this study is to refine the Grossman model by offering a plausible extension, as motivated by empirical observations, that increases its usability. Grossman outlined multiple factors affecting a person's tendency to invest in health capital, but apprehension on a societal level was not one of them. Given the observation that anxious people are more prone to seek emergency health services, one could postulate that the degree of apprehension on a societal level is associated with higher emergency department attendance.

On the basis of previous research in combination with Grossman's theoretical framework, we therefore first outline a general hypothesis, as follows:

> *Hypothesis A: An increased degree of societal apprehension is associated with an increased tendency to invest in health capital.*

Given the previously mentioned indices' representation of societal apprehension, the hypotheses we test in our dataset accumulates to:

> *Hypothesis 1: An increased use of apprehension-reflecting vocabulary in established news media is associated with increased emergency department visits.*
>
> *Hypothesis 2: An increased use of apprehension-reflecting vocabulary in social media is associated with increased emergency department visits.*

6. Data

In this section, we present the data used to test our hypotheses.

6.1.1 Emergency department attendance

Data regarding the number of daily patient visits at the hospitals included was collected from GVR (Gemensamt Vårdregister, Stockholm County Council's common health registry), and summarized for analysis (34). In total, data from the ten-year period 01.01.2006-12.31.2015 was collected. The data was complete without any missing values. All data was anonymized, with no possibility to identify individuals.

6.1.2 Apprehension-reflecting vocabulary in news media

Data on the quantitative use of apprehension-reflecting vocabulary in news media was collected from Retriever, a private company specializing in surveillance and analysis of media. The company is currently a member of International Association for Measurement and Evaluation of Communication, an international organization for companies offering evaluatory services within the field of communication (41).

The Retriever database consists of over 700 Swedish daily newspapers and similar news publications. Digital copies of the original articles are made available in the database the same day as they are made public. The company has the ability to manually search the database with a high degree of precision. Searches can be defined by source, date, and keywords. Several keywords can be included and typical commands such as AND, OR, or NEAR can be used. The search can also be made to include only parts of a word, in order to capture, for example, a word with different endings (41).

A search of the Retriever database was conducted covering every day for a ten-year period from 01.01.2006 to 12.31.2015 This resulted in an index fluctuating daily on the basis of the presence of a number of keywords in the selected Swedish news publications. The publications were chosen based on them having a substantial user base in Stockholm County. A list of selected keyword and news publications can be found in Appendix A.

6.1.3 Apprehension-reflecting vocabulary in social media

The increasing availability of text data in open sources such as social media has created new opportunities for research based on large-quantity text analysis (42). In fact, the term *buzz monitoring* has been established and refers to the act of monitoring text sources for mentions of a specific textual content, such as a product or a brand. It has previously been established that word-of-mouth widely affects consumer behavior and brand reputation. The idea is that the words used in a document can provide us with useful and reliable information indicative of its content and sentiment (43, 44). As a result of this development, Gavagai, a private company specializing in analysis of large quantities of text, was founded (45). Briefly, the company analyzes the sentiments expressed in selected social medias platforms and news outlets by calculating the frequency with which a specific word and its associated synonyms (which collectively represent a sentiment) is used. Indeed, sentiment analysis of news texts has been investigated and found to possess some predictive capacity for stock price changes in several recent publications (44, 46). Several other applications of sentiment analysis have also been proposed, including tracking public emotion to foresee security threats or disorderly public behavior as well as identification and surveillance of certain individuals. Gavagai's services has previously been used by clients for mainly trademark surveillance. Although the system has not been validated externally, psychologists were consulted during the formation of the search and recognition algorithms. Among their clients are consulting firms, asset management corporations, and governmental agencies (47). Technically, the system is self-learning, as it observes words that are commonly used together with the specific word chosen for analysis. However, any manual broadening of its search to include several words (for example, angry and frustrated may both express the same sentiment) must be manually cleared.

To create the social media apprehension index, the Gavagai tracker registered one observation every time the word *'I'* appeared in combination with any of the selected words related to apprehension. In total, data was collected daily during the period 01.25.2013 to 12.31.2015. Considering that the combination of *'I'* and an apprehension-related word was necessary to register an observation, our data mostly represent the sentiment expressed in social media, where *'I'* is far more commonly used in sentences as compared to in established news media (48). Drawbacks of the system includes the inability to recognize sarcasm, irony, or negations. For instance, the tweet *'I have never been afraid'* would be counted as one observation of apprehension and included in the assessment of that sentiment. However, the frequency of irony is arguably limited as words are deemed to be most commonly used to express their inherent meaning (48). Due to a non-disclosure agreement, we are not able to present a list of keywords and medias used in the index.

7. Empirical method

In order to assess the proposed effect of apprehension on emergency department attendance, an ordinary least squares (OLS) regression analysis was performed with ED attendance constituting the dependent variable. Based on largely accordant previous research, calendar (weekday and month) and day-after-holiday variables were included as control dummy variables. Different day-after-holiday variables were created depending on the length of the holiday, as it was deemed reasonable to believe that the length of the holiday correlates with the increase in attendance the next working day. Monday and December were chosen as the weekday and month of reference respectively. Population increases was adjusted for by including a variable for the yearly population in Stockholm County Council. Over the study period, there was approximately an 18 % increase in the amount of inhabitants, which undoubtedly influences the number of emergency department visits. Population data was collected from Statistics Sweden (49). Population increases were assumed to increase continuously, as inhabitant data was only available for December 31 each year. Further, time itself was included as an independent variable as empirical observations from emergency departments suggest increased patient visits in excess of the expected increases resulting from a growing population (12). A Breusch-Pagan test was performed to assess for heteroscedasticity and robust standard errors used accordingly. A Durbin-Watson statistic was calculated to test for autocorrelation. Correlation measurements were calculated using Spearman's rho since variables were not normally distributed, as assessed by Kolmogorov-Smirnov's test. Finally, an assessment of multicollinearity was done. The result of this test can be seen in Appendix B.

In order to investigate whether apprehension on a specific day lingered on to the next upcoming days, lag effects were included. A total amount of seven lag effects were deemed to be reasonable and therefore included (as the effect of apprehension was assumed to linger on a total of seven days). Also, lead effects were examined. Lead effects test whether apprehension on a particular day predicts attendance on the previous day. All statistical analysis was performed in STATA 13 and a two-tailed $\alpha<0.05$ was considered statistically significant.

The final regressions thus becomes:

News media:

$$Y_t = \beta_0 + \beta_1 * t + \beta_2 * pop + \beta_3 * news + \beta_4 * news\_lag1 \ + \beta_5 * news\_lag2 + \beta_6 * news\_lag3 + \beta_7 * news\_lag4 + \beta_8 * news\_lag5 + \beta_9 * news\_lag6 + \beta_{10} * news\_lag7 + \beta_{11} * tue + \beta_{12} * wed + \beta_{13} * thu + \beta_{14} * fri + \beta_{15} * sat + \beta_{16} * sun + \beta_{17} * jan + \beta_{18} * feb + \beta_{19} * mar + \beta_{20} * apr + \beta_{21} * may + \beta_{22} * june + \beta_{23} * july + \beta_{24} * aug + \beta_{25} * sept + \beta_{26} * oct + \beta_{27} * nov + \beta_{28} * free1 + \beta_{29} * free2 + \beta_{30} * free3 + \beta_{31} * free4 + \varepsilon_t$$

Social media:

$$Y_t = \beta_0 + \beta_1 * t + \beta_2 * pop + \beta_3 * social + \beta_4 * social\_lag1 \ + \beta_5 * social\_lag2 + \beta_6 * social\_lag3 + \beta_7 * social\_lag4 + \beta_8 * social\_lag5 + \beta_9 * social\_lag6 + \beta_{10} * social\_lag7 + \beta_{11} * tue + \beta_{12} * wed + \beta_{13} * thu + \beta_{14} * fri + \beta_{15} * sat + \beta_{16} * sun + \beta_{17} * jan + \beta_{18} * feb + \beta_{19} * mar + \beta_{20} * apr + \beta_{21} * may + \beta_{22} * june + \beta_{23} * july + \beta_{24} * aug + \beta_{25} * sept + \beta_{26} * oct + \beta_{27} * nov + \beta_{28} * free1 + \beta_{29} * free2 + \beta_{30} * free3 + \beta_{31} * free4 + \varepsilon_t$$

Where:

TABLE 7 REGRESSION VARIABLES	
t = Time, one unit increase for every consecutive day pop = Population, continuously adjusted throughout the study period news = The quantity of apprehension-reflecting vocabulary in news media social = The quantity of apprehension-reflecting vocabulary in social media _lag_x = A variable depicting a lag effect of apprehension of x days tue = Tuesday wed = Wednesday thu = Thursday fri = Friday sat = Saturday	sun = Sunday jan = January feb = February mar = March apr = April jun = June jul = July aug = August sept = September oct = October nov = November free1 = The day after a one-day holiday free2 = The day after a two-day holiday free3 = The day after a three-day holiday free4 = The day after a four-day holiday

Table 7: List of regression variables.

8. Results

The following section is dedicated to presenting our results.

8.1 Descriptive statistics

TABLE 8.1 DESCRIPTIVE STATISTICS					
	Mean (95 % CI)	**Median (range)**	**Std. dev.**	**Missing values**	**Time period (daily obs.)**
ED attendance	1687 (1680-1695)	1678 (932-2514)	225	None	01.01.2006-12.31.2015
News media index	603 (600-607)	597 (210-1608)	108	None	01.01.2006-12.31.2015
Social media index	2906 (2849-2964)	2790 (1063-7426)	958	36	01.25.2013-12.31.2015

Table 8.1: Descriptive statistics for the variables ED attendance, news media, and social media. Obs. = observations.

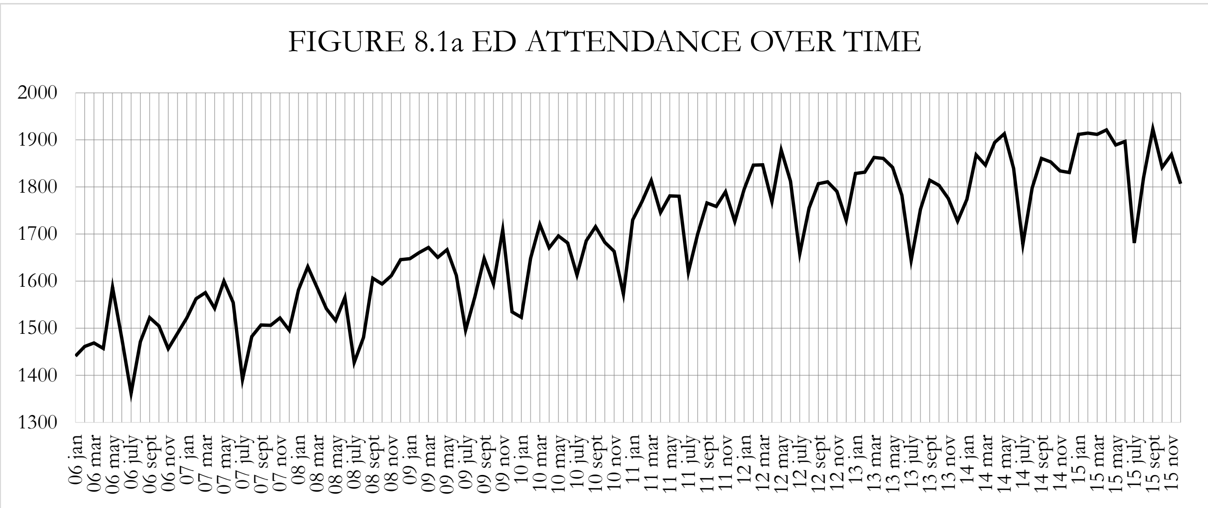

Figure 8.1a: Number of daily ED visits, averaged per month over time.

ED attendance displayed linear increase over time with a simple regression coefficient of 0.121. Seasonal trends were noted with overall lower attendance during summer months and higher during spring months. Distinguishably low attendances were noted around the Christmas and Midsummer holidays each year.

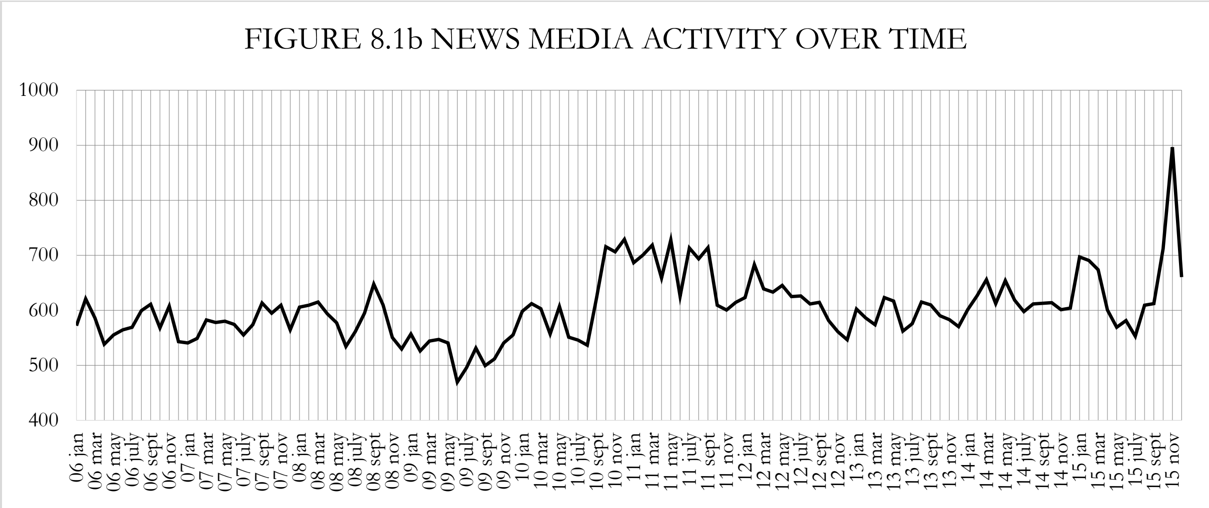

Figure 8.1b: Daily value of the news media index, averaged per month over time.

The frequency of apprehension-reflecting vocabulary used in news media remained relatively constant throughout the study period, although a few exceptions were observed. Fewer apprehensive words were used between January 2009 and November 2009, while higher quantities were expressed in November 2010 up until September 2011. A remarkably high activity was noted during November 2015, likely explained to a large extent by the highly publicized Paris attacks that took place on November 13, 2015.

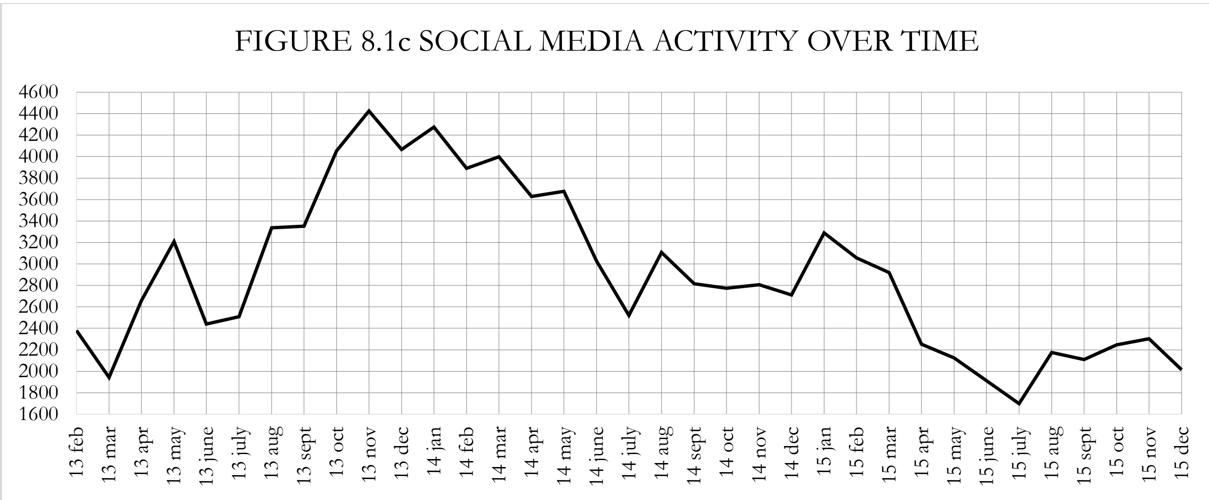

Figure 8.1c: Daily value of the social media index, averaged per month over time.

The social media index variable had 36 missing values. The index was found to fluctuate with an average of 1.8 local valleys per week as defined by a lower value than both the previous and consecutive day (which is not apparent in Figure 8.1c as it is based on monthly average values).

Therefore, missing values were filled in using an average of the three days prior to the missing value. A higher volatility of the index was noted prior to December 2013.

8.2 Correlation between the two measurements of apprehension

TABLE 8.2 CORRELATION BETWEEN INDICES								
Same-day correlation			**Corr. between social media index and news media index of subsequent days**			**Corr. between news media index and social media index of subsequent days**		
		social			social			news
news	corr.	0.108***	news_lead1	corr.	0.102***	social_lead1	corr.	0.023
			news_lead2	corr.	0.093***	social_lead2	corr.	0.009
			news_lead3	corr.	0.083***	social_lead3	corr.	0.031
			news_lead4	corr.	0.053*	social_lead4	corr.	0.042
			news_lead5	corr.	0.017	social_lead5	corr.	0.051*
			news_lead6	corr.	0.020	social_lead6	corr.	0.057*
			news_lead7	corr.	0.094***	social_lead7	corr.	0.036

Table 8.2: Results from tests of correlation between the two indices. Corr. = correlation coefficient. Coefficients are marked differently depending on their level of significance. *** p<0.01, ** p<0.05, * p<0.1.

The use of apprehension-reflecting vocabulary in predominantly social media was significantly, but only weakly, correlated with its same-day equivalent in news media. Further, the social media index also correlated significantly with the news media index on the three subsequent days as well as on the seventh day. However, the reverse effect–news media index being correlated with ulterior-day social media index–was not observable.

8.3 The association between activity in news media and ED attendance

TABLE 8.3 REGRESSION WITH NEWS MEDIA

t	0.121***	thu	-172.7***	aug	11.38
	(0.00163)		(13.49)		(9.900)
news	0.301***	fri	-170.7***	sept	70.93***
	(0.0353)		(14.08)		(9.586)
news _lag1	-0.0675**	sat	-377.7***	oct	44.83***
	(0.0291)		(13.30)		(9.857)
news _lag2	-0.0481*	sun	-372.9***	nov	47.77***
	(0.0257)		(12.33)		(9.993)
news _lag3	0.0243	jan	54.89***	free1	126.7***
	(0.0279)		(10.67)		(28.65)
news _lag4	-0.0598**	feb	98.38***	free2	64.23***
	(0.0258)		(10.46)		(11.24)
news _lag5	-0.0100	mar	106.7***	free3	105.6***
	(0.0259)		(9.944)		(32.53)
news _lag6	0.0449*	apr	75.30***	free4	174.5***
	(0.0265)		(10.77)		(23.54)
news _lag7	-0.0751***	may	102.9***	Constant	1,533***
	(0.0241)		(10.59)		(22.49)
tue	-100.2***	june	67.52***		
	(13.73)		(11.29)		
wed	-160.4***	july	-79.12***		
	(13.37)		(10.35)		
Observations 3,645					
R-squared 0.817					
Adj. R-squared 0.816					

Table 8.3: Result of regressing ED attendance on time, the news media index, lag effects, and calendar variables. Beta coefficients are marked differently depending on their level of significance. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Most of the included predictory variables were found to be significant in the regression. Overall, 81.6 % of the variation in emergency department attendance could be explained by the included variables.

8.3.1 Calendar variables and time–the news media index

The results show that emergency department attendance decreased throughout the week and most prominently on weekends as compared to the attendance on Mondays. The results also show fewer patient visits in July but more all other months compared to December. The day-after-holiday variables were all also significantly associated with higher ED attendance.

Time was found to correlate strongly (Spearman's rho = 0.9997, $p<0.01$) with population, and consequently the population variable was excluded from the final regression to decrease multicollinearity. The beta coefficient for time of 0.121 was significant.

8.3.2 Apprehension and lag effects–the news media index

The amount of apprehension-reflecting vocabulary in news media was found to independently predict ED attendance with a beta coefficient of 0.301. Lag effects for day 1, 4 and 7 respectively were significant and associated with a small decrease in the amount of patient visits. Coefficients for the other lag variables were not significant.

8.3.3 Effect of introducing lead variables–the news media index

Lead variables were introduced to test whether apprehension on a particular day can predict attendance on the previous day. The introduction of lead variables did not change the direction or significance of the previously analyzed coefficients, as can be seen in Appendix C. The one-day lead variable was not significant unless additional-day lead variables were included as well. When further lead variables were included, the one-day lead variable had a small positive impact on attendance the previous day, approximately one fifth as compared to the effect of same-day apprehension. When two lead effects were included, the second-day lead variable had a negative significant coefficient. When including a third lead effect, the third-day variable was found to have small negative impact on attendance, while the significance of the second-day lead effect vanished. No more than two lead variables were simultaneously significant, even after including up to a seven-day lead variable.

8.4 The association between activity in social media and ED attendance

TABLE 8.4 REGRESSION WITH SOCIAL MEDIA

t	0.108***	thu	17.68	aug	28.41
	(0.0110)		(74.35)		(19.62)
social	0.0184**	fri	39.13	sept	87.23***
	(0.00787)		(74.45)		(19.00)
social _lag1	0.00125	sat	-217.9***	oct	55.95***
	(0.0104)		(74.27)		(19.14)
social _lag2	-0.00514	sun	-200.7***	nov	51.51***
	(0.00569)		(74.43)		(18.69)
social _lag3	-0.00125	jan	75.34***	free1	161.5***
	(0.00714)		(22.82)		(37.76)
social _lag4	-0.00580	feb	122.2***	free2	277.4***
	(0.00832)		(20.31)		(74.04)
social _lag5	0.00546	mar	125.3***	free3	190.0*
	(0.00511)		(19.43)		(100.6)
social _lag6	-0.00778	apr	129.1***	free4	251.5***
	(0.00587)		(21.58)		(26.29)
social _lag7	-0.00154	may	125.8***	Constant	1,413***
	(0.00491)		(20.75)		(85.29)
tue	89.78	june	78.41***		
	(73.36)		(22.12)		
wed	35.74	july	-101.5***		
	(74.31)		(21.02)		
Observations 1,064					
R-squared 0.773					
Adj. R-squared 0.766					

Table 8.4: Result of regressing ED attendance on time, the social media index, lag effects, and calendar variables. Beta coefficients are marked differently depending on their level of significance. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

When evaluating the effect of apprehension-reflecting vocabulary in predominantly social media on ED attendance, both similarities and differences with the previous regression were noted. Overall, a slightly lower adjusted R-squared was computed at 0.766. Coefficients generally had the same direction but fewer were found to be significant.

8.4.1 Calendar variables and time–the social media index

In contrast to the previous regression, weekdays were associated with higher attendance as compared the reference day (Monday). However, coefficients for weekdays were not statistically significant. Saturday and Sunday were again associated with much fewer patient visits. Month variables displayed the same pattern as the previous regression with higher attendance for all months except July as compared to the reference month (December). Time was noted to positively influence attendance with a coefficient of 0.108.

8.4.2 Apprehension and lag effects–the social media index

Societal apprehension as measured by the quantity of apprehension-reflecting vocabulary in social media was also found to significantly influence ED attendance, however with a substantially smaller coefficient compared to the previous regression (0.0184 versus 0.301 for news media). None of the lag effects were significant.

8.4.3 Effect of introducing a lead variable–the social media index

Introducing lead variables up to a total of seven days did not change the direction or significance of other coefficients in the regression. When including lead effects for one and two days, the two-day lead effect was significant. The significance of this coefficient vanished after including more lead variables, as can be seen in Appendix D.

9. Discussion

In this section, we discuss our findings in relation to the Grossman model and provide theoretical explanations for the results. We also discuss limitations of our study.

9.1 Findings from regression analysis

In this essay, we explored a hypothesized extension of the Grossman model as motivated both from a theoretical and empirical point of view. The extension can be seen as such in the sense that it proposes yet another explanation as to why and to what extent persons invest in health capital. This explanation, societal apprehension, has not previously been explored. We have also suggested and tested a way to empirically measure this source of investment. Our results demonstrate that apprehension independently predicts emergency department attendance when controlling for other established factors known to influence the amount of patient visits. A remarkably high adjusted R-squared of 0.816 (news media) and 0.766 (social media), is found. Notably, the inclusion of a variable for time significantly increased the adjusted R-squared from 0.53 (news media) and 0.747 (social media) to the current levels. Previous research regarding the determinants of ED attendance have seldom utilized long-term data of this magnitude and the effect of time as an independent variable. As such, the longevity of our data and the inclusion of the time variable likely explains the high adjusted R-squared in our study in comparison with previous studies. In accordance with this, the incremental increase of the adjusted R-squared when adding time as an explanatory variable was considerably higher for the ten-year news index than the three-year social media index. Furthermore, it is worth noting that the relationship between ED attendance and its determinants is arguably largely different as compared to many other relationships investigated using econometric analysis. Variations in ED attendance is predominantly determined by known factors, such as calendar variables and time, whereas the dependent variable in other econometric analysis debatably often is determined by a complex array of variables, often subjective in nature, that interact with one another in an unknown way. Therefore, obtaining a higher adjusted R-squared than in most econometric analysis can be expected in this study.

Interestingly, our results remain significant regardless of whether apprehension was assessed through established news media or social media, despite the weak correlation between them. The beta coefficients for apprehension as measured by news media and social media were 0.301 and 0.0184. Thus, as ten more apprehension-reflecting words are expressed, three and 0.2 additional patients can be expected to attend the ED, depending on the variable used. In comparison, the news media index and the social media index differs from day to day on average by 75 points and 444 points respectively. Therefore, on a typical day, a change of 23 and 8 patients respectively can be expected in the ED solely due to an average day-to-day change in the degree of societal apprehension. Naturally, these numbers vary over time and a graphical representation of this can be found in Appendix E. The average attendance over the study period was 1687, implying that 1.4 % and 0.5 % of patient visits per day can be explained by societal apprehension, depending on the variable used.

It is apparent from the above that apprehension indeed significantly and substantially influences ED attendance. In light of this, we can reject our two null hypotheses implied in section five, namely that an increased use of apprehension-reflecting vocabulary in established news media is not associated with increased emergency department visits, and that an increased use of apprehension-reflecting vocabulary in social media is not associated with increased emergency department visits.

Regarding calendar variables and general determinants of emergency department attendance, our results are in line with previous research (15, 19-27). We demonstrate a high average attendance on Mondays, which declines steadily throughout the week and bottoms out on Sundays. Moreover, the day after a holiday is associated with a significantly increased expected attendance, in accordance with findings from previous authors (19, 21). We find a lower attendance in July compared to all other months, which is likely explained by the fact the July is the most common holiday month in our study setting.

In our material, the variable based on news media had a markedly higher influence on ED attendance as compared to the variable based on social media. Overall, the qualitative content of the former variable was considerably better, as it covered a ten-year period versus only approximately three for the social media variable. Whereas the news media based index fluctuated throughout the entire study period around a mean of approximately 600 apprehensive-reflecting words per day, the social media based index did not exhibit any common baseline value and varied considerably over the years observed. Further, the news media variable did not have any missing values, compared to 36 for the latter. Also, the reliability of the latter variable may be questioned on the basis of a significantly higher volatility for the first year of observations without any apparent explanation. Therefore, the quality of data may be one factor explaining the different results observed with the two measurements of apprehension. It may also be the case that the apprehension on a societal level is simply better reflected in news media. Social media measurements are subject to abundant noise apart from actual apprehension: news media almost exclusively uses words in their inherent context, while words are used more freely in social media. The construction of the variables likely contributed to this. While the news media index was based on twenty carefully chosen words, 89 preselected words were included in Gavagai's social media index. As such, the precision with which the variable reflects true apprehension could be smaller. On the other hand, an advantage of the broader search is an increased possibility of detecting low-grade societal changes in apprehension. Indeed, the social media index correlated with future-day news index values, while the opposite was not observed. This indicates that social media intercepted changes in apprehension at an early stage, with subsequent coverage in established news outlets the following days. Finally, an additional explanation as to why the news media index has a stronger association to ED attendance compared to the social media index is that the while the former would reflect apprehension of society in general, the latter might be biased as it is used to a larger extent by younger persons. As young people are overrepresented in social media, and because young people are healthier than the average person, a higher magnitude of societal apprehension would be needed to push them over the threshold required to make them visit the ED.

When regressing using the news media index, we demonstrate a small negative lag effect of apprehension after one, four and seven days. In other words, an increased degree of apprehension on a particular day is independently associated with an increased attendance on the same day, but also a slightly lower attendance one, four, and seven days after respectively. One explanation for this might be that apprehension constitutes the trigger that makes people seek medical help for symptoms that they have previously neglected. Once medical help has been sought and received, the demand for health in the next upcoming days decreases. As such, apprehension decreases attendance rates in the next upcoming days. In theoretical terms, the individual may have replenished his health stock to an optimal level by visiting a physician, and thus does not need to provide further investment inputs in the near future. It should be noted that the magnitude of the demonstrated lag effects on attendance is small, only about one fifth of the effect of apprehension on same day attendance. No lag effect was demonstrated using the social media variable.

Significant lead effects were observed in both regressions. The presence of lead effects may indicate reverse causation or omitted variables, and thus slightly impedes our ability to draw causal conclusions. Although the degree of apprehension as measured by each index is highly correlated between two consecutive days (a Spearman's rho of 0.494 and 0.810 for news and social media respectively), this is controlled for when testing for lead effects, and thus does not explain the significance of some lead coefficients. A concrete and thinkable example of reverse causation would be an abnormally high emergency attendance on a particular day due to a flu season. As a result, discussions on social media platforms arise and reports in news media are published during the subsequent days. Another explanation could be a measurement error in the apprehension variable. As such, the variable may not only reflect the level of apprehension, but could also be subject to some random noise. Regarding omitted variable bias, there could hypothetically be a variable X that is associated with attendance on a distinct day and apprehension some days after. We cannot think of any variable that could satisfy this condition and thus deem it unlikely. Further, as our choice of independent variable of investigation was plausible from an empirical point of view, we deem the risk of confounding to be low.

It is worth stating the logical relationship between lag and lead effects: a lag effect with a negative coefficient has the same implication as a lead effect with a positive coefficient, if this is being measured the next day. For example, an apprehension-inducing event on a Monday might lead to lower ED attendance on Tuesday (negative lag effect) as people are pushed past a threshold and replenish their health capital stock on Monday, while media naturally reports relatively more the following day (Tuesday). This same event would, seen through the perspective of Tuesday, equal a positive lead effect: more people would visit the ED the previous day (Monday) given today's (Tuesday's) media reports.

9.2 Theoretical explanations

Our results demonstrate that the degree of societal apprehension independently influences the tendency to invest in health capital, as measured by the quantity of patient visits in the ED. In theoretical terms, apprehension may decrease the cost of capital as assessed by the individual. Basically, health capital is valued more highly among apprehensive individuals. Consequently, the optimal health stock shifts outwards as depicted in Figure 2.2, and additional inputs are supplied in order to obtain the new optimal health stock. Furthermore, a link can be made between our findings and the discussion of an uncertainty amendment to the Grossman model that Chang and Hren found reasonable (3, 4). Uncertainty regarding one's future health status may in itself be a cause of apprehension, and thus a logical link exists between the two.

An additional explanation could be that apprehension causes the MEI-curve to shift outward. This would imply that the presence of apprehension has a similar effect compared to education on health investment behavior: it causes the individual to become a more efficient utilizer of health capital, and hence a higher health stock is desired. In practical terms, it is easy to imagine a worried individual becoming more compliant in terms of medication and following instructions from healthcare professionals. As such, the apprehensiveness arguably entails an increased implementation of health-promoting measures given the same amount of healthcare expenditure. However, in comparison with education, the level of apprehension fluctuates continuously and as such, its effect is temporary and it is dynamic in its nature.

Van der Pol et al. investigated health investments measured in the form of adherence to physician advice, and found that they were affected by time preferences. The authors explained the findings as a result of individuals attaching disproportionately higher significance to outcomes that are closer in time (50). This may be linked to the former explanation proposed here: an increased state of apprehension may influence the valuation of present health stock capital in comparison to its future equivalent. Consequently, health investment tendencies increase in the present, resulting in higher emergency department attendances.

It may also be of importance to comment on what consequences our research has for the Grossman model. The finding that apprehension influences the mechanisms for investment in health could have implications for certain future usage of the model but is not necessarily to be taken into consideration by all. In particular, any usage involving the concept of an optimal health stock should take the effect of apprehension into account. However, the underlying assumptions of the model are in no way affected or changed by the results or implications of this research. As such, our extension is concordant with the underlying premises of the model.

9.3 Limitations of this study

Our data is subject to positive autocorrelation, as deemed by a Durbin-Watson statistic of 1.2. Although coefficients remain unbiased in the presence of autocorrelation, standard errors are underestimated, potentially resulting in variables appearing significant when they should not.

In order to reject the general null hypothesis implied in section 5, namely that an increased degree of societal apprehension is not associated with an increased tendency to invest in health capital, two proxies must be of satisfactory quality: ED attendance must be an appropriate proxy for investments in health capital and apprehension-reflecting vocabulary in media must be an appropriate proxy for societal apprehension. Physician's visits and medical care is commonly recognized as a type of health capital investment (29). Visiting the ED is likely the easiest way to meet a physician quickly, as primary care facilities often require scheduling an appointment at least several days prior to visiting. Taken together, it thus seems as though ED visits should not be challenged as a proxy for investments in health capital. Of course, there are potential drawbacks of this measurement of health capital investments. For example, the behavior of visiting an ED due to an increased level of apprehension may not be representative of all groups of people. This would imply the presence of selection bias in our material. Also, physician visits only constitute one way of investing in health capital. Our study has thus not evaluated the effect of apprehension on other investment methods.

When discussing whether apprehension-reflecting vocabulary in media is an appropriate proxy for societal apprehension, there are two parts of the proxy to consider. Firstly, the quality of the data itself. The social media based index had a total of 36 missing values which were replaced using an averaging technique as previously described. Also, the way data was recorded changed slightly during the course of measurement and a few additional apprehension-reflecting words were added over time by Gavagai. One might also question the non-constant trend the data displays over time, with increased volatility before December 2013. In comparison, the news media index was recorded from a single search of an archive of a constant amount of news articles. The index displays data over a ten-year period. Naturally, there were no missing values and a certain list of search keywords was chosen for the one time search. The monthly average of index values over time displays a non-increasing trend with a relatively constant volatility, with the exception of November 2015. Clearly, the social media index has flaws in comparison to the news media index.

Secondly, of importance is how well apprehension-reflecting vocabulary measured in this way is representative of societal apprehension. To our knowledge, no consensus exists regarding the optimal way of measuring the degree of societal apprehension, and no previous research is available regarding the validity of our approach. Being the first of its kind, our study had to use new techniques to answer the research question. We postulated that activity in media is reflective of the general level of apprehension in society. However, no data exists regarding which vocabulary most efficiently captures apprehension, and we had to rely on logic reasoning when formulating the content of the news media indices. In contrast, the social media index was completely designed by Gavagai.

In addition to the two arguments presented above, one might also question the appropriateness of the social media index as a proxy because of its issues with transparency. As previously mentioned, in accordance with the preference of the Gavagai founder, we are not able to disclose the list of 89 words used to created the index. It is worth noting that this fact became apparent to us only after our analysis was completed. Instead of excluding the social media index from our thesis, we chose to make use of it while clearly stating its limitations, as it contributes to answer our research question together with the superior news media index. Notably, the association of apprehension to both indices is significant and has the same direction, thereby reinforcing the likelihood of valid data and non-coincidental results. Also, in order replicate our results, it is indeed possible to contract Gavagai and obtain the same list of words.

Another potential explanation of our results worth mentioning is that media may influence rather than simply reflect an underlying level of societal apprehension. Indeed, several authors have demonstrated that mass media can directly impact the level of apprehension and health-seeking behavior among audiences (18, 36, 37, 39, 40). Naturally, media may reflect as well as influence audiences at the same time. Our research does not address to which extent this occurs.

Furthermore, yet another explanation to why we observe a significant effect of the indices on ED attendance could be that the indices capture the effect of another variable not controlled for. However, it is difficult to assess what variable this might be. For example, seasonal periods of increased sickness might explain deviations in ED attendance. During such periods, media's coverage of these epidemics using words such as sickness and death unquestionably increases. In order to test whether it was solely due to this effect that the apprehension-coefficient became significant, we excluded all words containing the partials sick and/or death from our analysis. After doing this, the indices were still significantly influential of ED attendance.

An alternative approach to our choice of OLS regression would have been an autoregressive model such as autoregressive integrated moving average (ARIMA) or seasonal autoregressive integrated moving average (SARIMA). Although such an approach would have been feasible, standard OLS regressions have been frequently used in previous research regarding ED attendance, and are more easily interpreted. Furthermore, a large study investigating the relative accuracy of different statistical methods used for predicting ED attendance found most models accounting for autocorrelation to provide only smaller improvements over a multiple linear regression. The authors concluded that multiple linear regression thus is a sensible approach for these purposes (23).

Finally, in regression analysis, one must often make a tradeoff between usability and complexity of a model. In our analysis, it might have been beneficial to include additional independent variables.

When reviewing previous research, it became clear that there is no consensus regarding which variables that should be included in excess of the ones analyzed in this paper to correctly account for ED attendance. However, universal consensus should not be expected as there are inevitably regional differences regarding the impact of different variables. For example, rainy days have been implicated as a factor decreasing attendance in certain studies (15, 20). However, the impact of rain is considerably higher in a region where most days are sunny, as opposed to a rainy region. As such, all in all, despite not including other variables that might have an influence on ED attendance, we believe that our analysis thoroughly accounts for the factors with a major impact on ED attendance.

10. Concluding remarks

To our knowledge, these results are the first to demonstrate an association between apprehension on a societal level and investments in health capital. We conclude that the degree of apprehension on a societal level influences investments in health capital through a change of the perceived cost of capital and/or a shift in the MEI-curve. Therefore, apprehension should be acknowledged as a factor influencing the optimal health stock of an individual. In contrast with previously established factors however, the effect of apprehension is temporary and continuously fluctuating. Overall, this study contributes to the continuous improvement of the most prominent and utilized model in the field of health economics.

10.1 Future research

Considering that this study pioneers in evaluating the role of apprehension in determining an individual's optimal health stock, more research is required to confirm our findings. Future researchers should attempt to replicate our findings in different study settings and using different proxies for health investments and apprehension respectively.

Future studies should also investigate the differential impact of apprehension on different patient populations. Empirically, non-urgent patients are those who stand for most of the fluctuation in daily attendance, and perhaps their health-related behavior is the most sensitive to changes in the degree of apprehension. A future study could therefore do well in restricting ED attendance data to include only persons arriving by own initiative (excluding persons arriving by referral from another physician or by ambulatory care), which would focus the analysis on this patient group. For the same reason, data could be restricted to include only persons categorized as least urgent in emergency departments' triage system. Of interest would also be to include only psychiatric emergency departments in a future study, the demand of which may be inherently different from somatic emergency departments.

.

11. References

(1) Folland, S., Goodman, A.C. and Stano, M. (2016). *The economics of health and health care*. Seventh edition. New York: Routledge.

(2) Grossman, M. (1972). On the concept of health capital and the demand for health. *Journal of Political Economy*, 80(2), pp. 223-255.

(3) Chang, F. (1996). Uncertainty and investment in health. *Journal of Health Economics*, 15(3), pp. 369–376.

(4) Hren, R. (2012). Theoretical shortcomings of the Grossman model. *Bulletin: Economics, Organisation and Informatics in Healthcare*, 28(1), pp. 63–75.

(5) Wagstaff, A. (1986). The demand for health: some new empirical evidence. *Journal of Health Economics*, 5(3), pp. 195–233.

(6) Asano, T. and Shibata, A. (2011). Risk and uncertainty in health investment. *European Journal of Health Economics*, 12(1), pp. 79–85.

(7) Michael, R.T. (2004). Sexual capital: an extension of Grossman's concept of health capital. *Journal of Health Economics*, 23(4), pp. 643–652.

(8) Bolin, K., Jacobson, L. and Lindgren, B. (2002). The family as the health producer–when spouses act strategically. *Journal of Health Economics,* 21(3), pp. 475–495.

(9) Backman, A.S., Blomqvist, P., Lagerlund, M., Carlsson-Holm, E. and Adami, J. (2008). Characteristics of non-urgent patients. Cross-sectional study of emergency department and primary care patients. *Scandinavian Journal of Primary Health Care,* 26(3), pp. 181–187.

(10) 1177 Vårdguiden, (2016). *Patientavgifter i Stockholms län*. [online] Available at: http://www.1177.se/Stockholm/Regler-och-rattigheter/Patientavgifter-i-Stockholm/ [Accessed February 29, 2016].

(11) The Södersjukhuset web site, (2016). *Verksamhetsområde akut.* [online] Available at: http://www.sodersjukhuset.se/Om-SOS/Organisation/Verksamhetsomraden/Akut/ [Accessed February 29, 2016].

(12) Miyan, M. and Nordberg, M. (2016). *Interview regarding implementations of the Grossman model and emergency departments.*

(13) Feinstein, J.S. (1993). The relationship between socioeconomic status and health: a review of the literature. *Milbank Quarterly*, 71(2), pp. 279–322.

(14) Galama, T.J. and van Kippersluis, H. (2013). Health inequalities through the lens of health capital theory: issues, solutions, and future directions. *Research on Economic Inequality*, 21, pp. 263–284.

(15) Kam, H.J., Sung, J.O. and Park, R.W. (2010). Prediction of daily patient numbers for a regional emergency medical center using time series analysis. *Healthcare Informatics Research*, 16(3), pp. 158–165.

(16) The American Hospital Association web site, (2016). *Trends affecting hospitals and health systems, April 2007*. [online] http://www.aha.org/research/reports/tw/chartbook/2007chartbook.shtml [Accessed Feb 17, 2016].

(17) Schweigler, L.M., Desmond, J.S., McCarthy, M.L., Bukowski, K.J., Ionides, E.L. and Younger, J.G. (2009). Forecasting models of emergency department crowding. *Academic Emergency Medicine: Official Journal of the Society for Academic Emergency Medicine*, 16(4), pp. 301–308.

(18) Filipkowski, K.B., Smyth, J.M., Rutchick, A.M., Santuzzi, A.M., Adya, M., Petrie, K.J. and Kaptein, A.A. (2010). Do healthy people worry? Modern health worries, subjective health complaints, perceived health, and health care utilization. *International Journal of Behavioral Medicine*, 17(3), pp. 182–188.

(19) Batal, H., Tench, J., McMillan, S., Adams, J. and Mehler, P.S. (2001). Predicting patient visits to an urgent care clinic using calendar variables. *Academic Emergency Medicine: Official Journal of the Society for Academic Emergency Medicine*, 8(1), pp. 48–53.

(20) Diehl, A.K., Morris, M.D. and Mannis, S.A. (1981). Use of calendar and weather data to predict walk-in attendance. *Southern Medical Journal*, 74(6), pp. 709–712.

(21) Holleman, D.R., Bowling, R.L. and Gathy, C. (1996). Predicting daily visits to a walk-in clinic and emergency department using calendar and weather data. *Journal of General Internal Medicine*, 11(4), pp. 237–239.

(22) Côté, M.J., Smith, M.A., Eitel, D.R. and Akçali, E. (2013). Forecasting emergency department arrivals: a tutorial for emergency department directors. *Hospital Topics*, 91(1), pp. 9–19.

(23) Jones, S.S., Thomas, A., Evans, R.S., Welch, S.J., Haug, P.J. and Snow, G.L. (2008). Forecasting daily patient volumes in the emergency department. *Academic Emergency Medicine: Official Journal of the Society for Academic Emergency Medicine*, 15(2), pp. 159–170.

(24) Wargon, M., Casalino, E. and Guidet, B. (2010). From model to forecasting: a multicenter study in emergency departments. *Academic Emergency Medicine: Official Journal of the Society for Academic Emergency Medicine*, 17(9), pp. 970–978.

(25) Sahu, S.K., Baffour, B., Harper, P.R. and Sarran, C. (2014). A hierarchical Bayesian model for improving short-term forecasting of hospital demand by including meteorological information. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 177(1), pp. 39–61.

(26) Wargon, M., Guidet, B., Hoang, T.D. and Hejblum, G. (2009). A systematic review of models for forecasting the number of emergency department visits. *Emergency Medicine Journal*, 26(6), pp. 395–399.

(27) Glass, R. and Friedman, D. (1977). Trends in the demand for emergency room services: the Mount Sinai Hospital. *Mount Sinai Journal of Medicine*, 44(4), pp. 560–565.

(28) Ekström, A., Kurland, L., Farrokhina, N., Castrén, M. and Nordberg, M. (2015). Forecasting emergency department visits using internet data. *Annals of Emergency Medicine*, 65(4), pp. 436–442.

(29) Redelmeier, D.A. and Vermeulen, M.J. (2011). Emergency department visits during an Olympic gold medal television broadcast. *Open Medicine: A Peer-Reviewed, Independent, Open-Access Journal*, 5(2), pp. 112–119.

(30) Grilli, R., Ramsay, C. and Minozzi, S. (2002). Mass media interventions: effects on health services utilisation. *Cochrane Database of Systematic Reviews*, (1).

(31) Noar, S.M. (2006). A 10-year retrospective of research in health mass media campaigns: where do we go from here? *Journal of Health Communication*, 11(1), pp. 21–42.

(32) Flowers, P., McDaid, L.M. and Knussen, C. (2013). Exposure and impact of a mass media campaign targeting sexual health amongst Scottish men who have sex with men: an outcome evaluation. *BioMed Central Public Health*.

(33) Scheufele, D.A. and Tewksbury, D. (2007). Framing, agenda setting, and priming: the evolution of three media effects models. *Journal of Communication*, 57(1), pp. 9–20.

(34) Kingdon, J.W. (2014). *Agendas, alternatives, and public policies*. Second edition. Harlow: Pearson.

(35) Kahneman, D. (2003). Maps of bounded rationality: psychology for behavioral economics. *The American Economic Review*, 93(5), pp. 1449–1475.

(36) Codish, S., Novack, L., Dreiher, J., Barski, L., Jotkowitz, A., Zeller, L. and Novack, V. (2014). Impact of mass media on public behavior and physicians: an ecological study of the H1N1 influenza pandemic. *Infection Control and Hospital Epidemiology*, 35(6), pp. 709–716.

(37) Nelissen, S., Beullens, K., Lemal, M. and Van den Bulck, J. (2015). Predictors of cancer fear: the association between mass media and fear of cancer among cancer diagnosed and nondiagnosed individuals. *Journal of Cancer Education: Official Journal of the American Association for Cancer Education*, 30(1), pp. 68–74.

(38) Nattinger, A.B., Hoffmann, R.G., Howell-Pelz, A. and Goodwin, J.S. (1998). Effect of Nancy Reagan's mastectomy on choice of surgery for breast cancer by US women. *Journal of the American Medical Association*, 279(10), pp. 762–766.

(39) Towers, S., Afzal, S., Bernal, G., Bliss, N., Brown, S., Espinoza, B., Jackson, J., Judson-Garcia, J., Khan, M., Lin, M., Mamada, R., Moreno, V.M., Nazari, F., Okuneye, K., Ross, M.L., Rodriquez, C., Medlock, J., Ebert, D. and Castillo-Chavez, C. (2015). Mass Media and the Contagion of Fear: The Case of Ebola in America. *Public Library of Science ONE*, 10(6).

(40) Leon, S.L., Cloutier, P. BéLair, M.A. and Cappelli, M. (2014). Media coverage of youth suicides and its impact on paediatric mental health emergency department presentations. *Healthcare Policy*, 10(1), pp. 97–107.

(41) Miyan, M. and Landin, P. (2016). *Interview regarding Retriever.*

(42) Gavagai, (2016). *The Gavagai living lexicon*. Stockholm: Gavagai.

(43) Gavagai, (2016). *Buzz monitoring in word space.* Stockholm: Gavagai.

(44) Gavagai, (2016), *Usefulness of sentiment analysis.* Stockholm: Gavagai.

(45) The Gavagai web site, (2016). [online] Available at: https://www.gavagai.se/ [Accessed March 18, 2016].

(46) Bollen, J., Mao, H. and Zeng, X.–J. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2 (1), pp. 1–8.

(47) The Gavagai web site, (2016). *Clients and partners.* [online] Available at: https://www.gavagai.se/clients-and-partners/ [Accessed March 18, 2016].

(48) Miyan, M. and Karlgren, J. (2016). *Interview regarding Gavagai*.

(49) The Statistics Sweden web site, (2016). *Befolkningsstatistik*. [online] Available from: http://www.scb.se/sv_/Hitta-statistik/Statistik-efter-amne/Befolkning/Befolkningens-sammansattning/Befolkningsstatistik/ [Accessed April 18, 2016].

(50) Van der Pol, M., Hennessy, D. and Manns, B. (2016). The role of time and risk preferences in adherence to physician advice on health behavior change. *European Journal of Health Economics*.

Appendix A–Keywords and publications of news media index

Keywords	*attack* OR *attack** OR **attack* *attentat* OR *attentat** OR **attentat* *bekymmer* OR *bekymmer** OR* *bekymmer* *chock* OR *chock** OR **chock* *dåd* OR *dåd** OR **dåd* *död* OR *död** OR **död* *fara* OR *fara** OR **fara* *hot* OR *hot** OR **hot* *kaos* OR *kaos** OR **kaos* *katastrof* OR *katastrof** OR **katastrof* *krig* OR *krig** OR **krig* *kris* OR *kris** OR **kris* *larm* OR *larm** OR **larm* *mord* OR *mord** OR **mord* *oro* OR *oro** OR **oro* *otrygg* OR *otrygg** OR **otrygg* *panik* OR *panik** OR **panik* *rädsla* OR *rädsla** OR **rädsla* *sjuk* OR *sjuk** OR **sjuk* *terror* OR *terror** OR **terror*	
News publications	**Name**	**Format**
	Svenska Dagbladet	Printed
	Dagens Nyheter	Printed
	Aftonbladet	Printed
	Expressen	Printed
	TT–Tidningarnas Telegrambyrå	Printed
	Sveriges Radio Ekot	Web
	SVT Eyheter	Web

Table A: A list of keyword and news publications used to create the news media index. The asterisk and the OR-command indicates that both words ending with and starting with the selected word were recognized in the search results.

Appendix B–Assessment of multicollinearity

TABLE B MULTICOLLINEARITY OF THE REGRESSIONS			
News media index regression		Social media index regression	
Variable	**VIF**	**Variable**	**VIF**
fri	5.44	sun	20.69
tue	5.30	fri	20.66
wed	5.25	thu	20.65
thu	5.22	sat	20.62
sat	5.05	wed	20.59
sun	4.73	tue	20.11
free2	4.01	free2	19.63
news_lag1	2.76	social_lag5	3.92
news_lag4	2.62	social_lag6	3.84
news_lag3	2.60	social_lag4	3.81
news_lag2	2.58	social_lag3	3.78
news_lag5	2.56	social_lag2	3.78
news_lag6	2.55	social_lag1	3.75
news	2.08	social_lag7	3.11
news_lag7	2.03	social	3.08
may	1.86	july	2.01
mar	1.85	june	1.93
july	1.85	mar	1.92
jan	1.85	apr	1.90
aug	1.84	may	1.90
oct	1.84	aug	1.86
apr	1.83	feb	1.84
june	1.82	sept	1.84
nov	1.82	oct	1.84
sept	1.82	nov	1.82
feb	1.79	jan	1.64
t	1.13	t	1.47
free1	1.09	free3	1.16
free4	1.08	free4	1.08
free3	1.05	free1	1.04
Mean VIF	2.64	Mean VIF	6.58

Table B: The results from a VIF-test in STATA13.

The presence of multicollinearity does not cause biased coefficient estimates, however it increases standard errors, thus making it more difficult to reject null hypotheses. According to a rule of thumb, variables with a VIF (variance inflation factor) larger than 10 are subject to concern. Because none of the variables possess this characteristic in the news media index regression, it is reasonable to conclude that multicollinearity should not affect our interpretations of the data. Regarding the social media index, several calendar variables have a VIF higher than 10. However, multicollinearity can safely be ignored when the variables presenting a high VIF are control variables rather than variables of interest, which is the case in this regression analysis.

Appendix C–Effect of introducing a lead variable–news media

TABLE C REGRESSIONS WITH NEWS INDEX AND LEAD VARIABLES

VARIABLES	(1) edatt	(2) edatt	(3) edatt	(4) edatt	(5) edatt	(6) edatt	(7) edatt	(8) edatt
t	0.121***	0.121***	0.121***	0.121***	0.122***	0.122***	0.122***	0.122***
	(0.00161)	(0.00161)	(0.00161)	(0.00162)	(0.00162)	(0.00162)	(0.00163)	(0.00163)
news	0.301***	0.282***	0.284***	0.286***	0.286***	0.286***	0.286***	0.287***
	(0.0214)	(0.0240)	(0.0240)	(0.0240)	(0.0240)	(0.0240)	(0.0240)	(0.0241)
news_lag1	-0.0675***	-0.0693***	-0.0672***	-0.0664***	-0.0649***	-0.0644***	-0.0653***	-0.0679***
	(0.0246)	(0.0245)	(0.0245)	(0.0245)	(0.0245)	(0.0246)	(0.0247)	(0.0247)
news_lag2	-0.0481**	-0.0484**	-0.0479**	-0.0462*	-0.0456*	-0.0465*	-0.0463*	-0.0455*
	(0.0238)	(0.0237)	(0.0237)	(0.0237)	(0.0237)	(0.0238)	(0.0238)	(0.0238)
news_lag3	0.0243	0.0255	0.0269	0.0293	0.0313	0.0311	0.0309	0.0300
	(0.0239)	(0.0238)	(0.0238)	(0.0238)	(0.0239)	(0.0239)	(0.0239)	(0.0239)
news_lag4	-0.0598**	-0.0614**	-0.0596**	-0.0559**	-0.0580**	-0.0582**	-0.0582**	-0.0575**
	(0.0239)	(0.0239)	(0.0239)	(0.0240)	(0.0240)	(0.0240)	(0.0240)	(0.0240)
news_lag5	-0.0100	-0.00852	-0.00516	-0.00820	-0.00744	-0.00761	-0.00752	-0.00603
	(0.0237)	(0.0236)	(0.0237)	(0.0237)	(0.0237)	(0.0237)	(0.0237)	(0.0237)
news_lag6	0.0449*	0.0389*	0.0365	0.0381	0.0373	0.0376	0.0374	0.0369
	(0.0236)	(0.0237)	(0.0237)	(0.0237)	(0.0237)	(0.0237)	(0.0237)	(0.0237)
news_lag7	-0.0751***	-0.0765***	-0.0748***	-0.0746***	-0.0736***	-0.0735***	-0.0736***	-0.0713***
	(0.0211)	(0.0210)	(0.0210)	(0.0210)	(0.0210)	(0.0210)	(0.0210)	(0.0211)
news_lead1		0.0330	0.0555**	0.0572**	0.0580**	0.0583**	0.0581**	0.0600**
		(0.0211)	(0.0238)	(0.0238)	(0.0238)	(0.0238)	(0.0238)	(0.0238)
news_lead2			-0.0431**	-0.0171	-0.0159	-0.0158	-0.0158	-0.0150
			(0.0210)	(0.0238)	(0.0238)	(0.0238)	(0.0238)	(0.0238)
news_lead3				-0.0499**	-0.0406*	-0.0407*	-0.0409*	-0.0401*
				(0.0210)	(0.0238)	(0.0238)	(0.0238)	(0.0238)
news_lead4					-0.0193	-0.0211	-0.0214	-0.0199
					(0.0210)	(0.0237)	(0.0238)	(0.0238)
news_lead5						0.00275	-0.000699	0.000261
						(0.0210)	(0.0238)	(0.0238)
news_lead6							0.00643	0.0214
							(0.0211)	(0.0238)
news_lead7								-0.0300
								(0.0211)
tue	-100.2***	-99.00***	-98.65***	-97.34***	-100.7***	-101.1***	-101.0***	-100.2***
	(10.49)	(10.48)	(10.47)	(10.48)	(10.83)	(10.97)	(10.97)	(10.98)
wed	-160.4***	-159.5***	-157.5***	-163.1***	-164.8***	-165.1***	-165.2***	-164.5***
	(10.44)	(10.42)	(10.45)	(10.70)	(10.74)	(10.83)	(10.84)	(10.84)
thu	-172.7***	-171.5***	-175.4***	-176.2***	-178.0***	-178.3***	-178.3***	-177.5***
	(10.42)	(10.39)	(10.56)	(10.55)	(10.61)	(10.77)	(10.77)	(10.78)
fri	-170.7***	-166.4***	-166.2***	-167.7***	-169.2***	-169.4***	-169.5***	-167.2***
	(10.64)	(10.93)	(10.93)	(10.93)	(10.96)	(11.07)	(11.08)	(11.13)
sat	-377.7***	-377.1***	-377.5***	-377.8***	-379.3***	-379.6***	-379.8***	-381.8***
	(10.26)	(10.23)	(10.23)	(10.22)	(10.26)	(10.42)	(10.47)	(10.57)
sun	-372.9***	-371.8***	-371.2***	-371.9***	-373.2***	-373.3***	-372.7***	-372.1***
	(9.909)	(9.910)	(9.912)	(9.909)	(9.936)	(10.25)	(10.41)	(10.42)
jan	54.89***	52.69***	53.14***	53.47***	53.09***	52.74***	52.47***	52.61***

	(7.852)	(7.846)	(7.855)	(7.861)	(7.876)	(7.892)	(7.926)	(7.944)
feb	98.38***	96.39***	96.83***	97.27***	97.11***	96.76***	96.55***	96.49***
	(7.981)	(7.971)	(7.979)	(7.985)	(7.993)	(8.006)	(8.027)	(8.038)
mar	106.7***	104.9***	105.1***	105.0***	104.7***	104.3***	104.2***	103.8***
	(7.784)	(7.770)	(7.774)	(7.775)	(7.778)	(7.786)	(7.798)	(7.801)
apr	75.30***	73.71***	73.73***	73.72***	73.35***	73.01***	72.81***	72.66***
	(7.846)	(7.828)	(7.832)	(7.834)	(7.839)	(7.850)	(7.870)	(7.877)
may	102.9***	100.8***	101.0***	101.1***	100.7***	100.3***	100.2***	99.83***
	(7.792)	(7.783)	(7.788)	(7.790)	(7.797)	(7.806)	(7.815)	(7.818)
june	67.52***	65.94***	65.74***	65.28***	64.62***	64.28***	64.12***	63.70***
	(7.838)	(7.820)	(7.824)	(7.824)	(7.830)	(7.838)	(7.852)	(7.855)
july	-79.12***	-80.86***	-81.07***	-81.45***	-81.92***	-82.25***	-82.39***	-82.80***
	(7.775)	(7.758)	(7.762)	(7.762)	(7.765)	(7.772)	(7.784)	(7.787)
aug	11.38	9.468	9.631	9.739	9.483	9.123	8.919	8.784
	(7.757)	(7.743)	(7.749)	(7.752)	(7.758)	(7.771)	(7.792)	(7.801)
sept	70.93***	69.00***	69.21***	69.36***	69.09***	68.75***	68.56***	68.36***
	(7.828)	(7.815)	(7.821)	(7.824)	(7.829)	(7.840)	(7.858)	(7.865)
oct	44.83***	42.98***	43.11***	43.03***	42.63***	42.29***	42.14***	41.76***
	(7.756)	(7.741)	(7.746)	(7.748)	(7.751)	(7.758)	(7.771)	(7.775)
nov	47.77***	45.68***	46.07***	46.48***	46.33***	45.97***	45.73***	45.76***
	(7.838)	(7.830)	(7.838)	(7.846)	(7.856)	(7.873)	(7.900)	(7.916)
free1	126.7***	125.1***	126.1***	126.0***	126.2***	126.2***	125.8***	125.3***
	(22.05)	(22.00)	(22.00)	(21.98)	(21.98)	(21.99)	(22.02)	(22.01)
free2	64.23***	64.89***	65.19***	64.41***	63.64***	63.59***	63.52***	63.86***
	(9.289)	(9.269)	(9.267)	(9.265)	(9.274)	(9.282)	(9.286)	(9.283)
free3	105.6***	106.0***	104.5***	105.7***	119.3***	119.3***	119.2***	119.1***
	(27.45)	(27.36)	(27.36)	(27.34)	(28.37)	(28.38)	(28.39)	(28.37)
free4	174.5***	174.0***	174.9***	175.7***	175.4***	175.3***	174.9***	175.9***
	(27.82)	(27.73)	(27.72)	(27.71)	(27.70)	(27.71)	(27.74)	(27.73)
Constant	1,533***	1,530***	1,536***	1,544***	1,549***	1,549***	1,549***	1,552***
	(16.94)	(17.24)	(17.45)	(17.72)	(18.07)	(18.10)	(18.24)	(18.34)
Observations	3,645	3,644	3,643	3,642	3,641	3,640	3,639	3,638
R-squared	0.817	0.819	0.819	0.819	0.819	0.819	0.819	0.820
Adj. R-squared	0.816	0.817	0.817	0.817	0.818	0.818	0.817	0.818

Table C: Result of regressing using the news media index while including lead variables, as discussed in section 8.3.3. Beta coefficients are marked differently depending on their level of significance. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Appendix D–Effect of introducing a lead variable–social media

TABLE D REGRESSIONS WITH SOCIAL MEDIA INDEX AND LEAD VARIABLES

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	edatt	edatt	edatt	edatt	edatt	edatt	edatt	edatt
t	0.108***	0.110***	0.108***	0.106***	0.105***	0.105***	0.105***	0.104***
	(0.0117)	(0.0116)	(0.0117)	(0.0117)	(0.0118)	(0.0118)	(0.0119)	(0.0119)
social	0.0184***	0.0167***	0.0185***	0.0188***	0.0188***	0.0200***	0.0201***	0.0198***
	(0.00541)	(0.00591)	(0.00596)	(0.00596)	(0.00599)	(0.00609)	(0.00612)	(0.00610)
social_lag1	0.00125	0.000704	0.00112	0.000865	0.00108	0.00102	0.00104	0.00104
	(0.00598)	(0.00598)	(0.00598)	(0.00599)	(0.00608)	(0.00610)	(0.00611)	(0.00610)
social_lag2	-0.00514	-0.00604	-0.00617	-0.00504	-0.00463	-0.00475	-0.00471	-0.00550
	(0.00600)	(0.00595)	(0.00596)	(0.00606)	(0.00608)	(0.00608)	(0.00609)	(0.00609)
social_lag3	-0.00125	-0.00315	-0.00203	-0.00152	-0.00147	-0.00143	-0.00145	-0.000940
	(0.00600)	(0.00596)	(0.00606)	(0.00607)	(0.00608)	(0.00608)	(0.00610)	(0.00609)
social_lag4	-0.00580	-0.000941	-0.000192	-0.000177	4.30e-05	-0.000272	-0.000235	-0.00101
	(0.00603)	(0.00608)	(0.00609)	(0.00609)	(0.00610)	(0.00611)	(0.00612)	(0.00612)
social_lag5	0.00546	0.00396	0.00406	0.00434	0.00405	0.00429	0.00428	0.00450
	(0.00612)	(0.00609)	(0.00608)	(0.00608)	(0.00609)	(0.00611)	(0.00613)	(0.00611)
social_lag6	-0.00778	-0.00862	-0.00809	-0.00869	-0.00824	-0.00851	-0.00848	-0.00833
	(0.00607)	(0.00601)	(0.00601)	(0.00602)	(0.00604)	(0.00605)	(0.00606)	(0.00605)
social_lag7	-0.00154	-0.00262	-0.00288	-0.00222	-0.00222	-0.00197	-0.00192	-0.00103
	(0.00546)	(0.00542)	(0.00542)	(0.00543)	(0.00544)	(0.00546)	(0.00548)	(0.00550)
social_lead1		0.00365	0.00888	0.0102*	0.0104*	0.0101*	0.00990	0.00980
		(0.00536)	(0.00590)	(0.00595)	(0.00596)	(0.00597)	(0.00608)	(0.00608)
social_lead2			-0.0113**	-0.00722	-0.00643	-0.00644	-0.00637	-0.00361
			(0.00535)	(0.00590)	(0.00596)	(0.00597)	(0.00599)	(0.00608)
social_lead3				-0.00881	-0.00632	-0.00592	-0.00587	-0.00676
				(0.00537)	(0.00593)	(0.00598)	(0.00600)	(0.00600)
social_lead4					-0.00535	-0.00399	-0.00389	-0.00389
					(0.00538)	(0.00595)	(0.00601)	(0.00600)
social_lead5						-0.00286	-0.00258	-0.00161
						(0.00541)	(0.00597)	(0.00601)
social_lead6							-0.000620	0.00260
							(0.00543)	(0.00595)
social_lead7								-0.00701
								(0.00545)
tue	89.78**	91.52**	90.37**	86.35**	84.41**	85.07**	85.10**	84.06**
	(37.93)	(37.59)	(37.55)	(37.63)	(38.64)	(38.67)	(38.71)	(38.63)
wed	35.74	40.03	36.92	30.01	30.84	31.82	31.80	30.82
	(38.38)	(38.05)	(38.04)	(38.23)	(39.17)	(39.24)	(39.28)	(39.20)
thu	17.68	23.88	17.48	15.02	16.65	17.43	17.37	17.19
	(38.43)	(38.15)	(38.21)	(38.22)	(39.20)	(39.26)	(39.29)	(39.23)
fri	39.13	41.93	41.25	40.09	41.33	42.18	42.01	41.48
	(38.45)	(38.24)	(38.19)	(38.18)	(39.16)	(39.20)	(39.25)	(39.23)
sat	-217.9***	-216.9***	-215.9***	-217.6***	-216.6***	-215.4***	-215.8***	-219.9***
	(38.41)	(38.06)	(38.01)	(38.00)	(38.99)	(39.03)	(39.11)	(39.16)
sun	-200.7***	-200.2***	-200.5***	-202.5***	-202.0***	-201.1***	-201.5***	-202.4***
	(38.47)	(38.12)	(38.07)	(38.06)	(38.89)	(38.91)	(39.07)	(38.99)
jan	75.34***	70.28***	70.94***	71.13***	71.59***	70.91***	71.11***	68.73***

	(16.20)	(16.10)	(16.12)	(16.15)	(16.21)	(16.25)	(16.30)	(16.30)
feb	122.2***	118.2***	116.7***	115.4***	115.0***	113.8***	113.9***	110.6***
	(14.90)	(14.81)	(14.82)	(14.85)	(14.88)	(14.92)	(14.98)	(15.00)
mar	125.3***	121.0***	119.7***	118.4***	118.1***	116.9***	117.0***	113.9***
	(14.54)	(14.44)	(14.46)	(14.49)	(14.53)	(14.57)	(14.63)	(14.65)
apr	129.1***	124.6***	123.4***	122.2***	121.9***	120.8***	120.9***	117.8***
	(14.67)	(14.57)	(14.59)	(14.62)	(14.69)	(14.74)	(14.79)	(14.81)
may	125.8***	121.1***	120.4***	119.3***	119.2***	118.0***	118.1***	114.8***
	(14.44)	(14.34)	(14.36)	(14.39)	(14.43)	(14.47)	(14.54)	(14.56)
june	78.41***	74.05***	71.56***	69.47***	68.69***	67.21***	67.29***	63.44***
	(14.77)	(14.69)	(14.73)	(14.78)	(14.87)	(14.93)	(15.01)	(15.06)
july	-101.5***	-106.1***	-108.1***	-109.7***	-110.2***	-111.5***	-111.4***	-115.1***
	(14.86)	(14.78)	(14.81)	(14.86)	(14.90)	(14.96)	(15.02)	(15.06)
aug	28.41**	23.69*	22.77	21.62	21.46	20.30	20.47	17.72
	(14.31)	(14.22)	(14.24)	(14.27)	(14.31)	(14.35)	(14.41)	(14.42)
sept	87.23***	82.24***	81.72***	81.15***	81.37***	80.39***	80.58***	77.75***
	(14.44)	(14.35)	(14.37)	(14.41)	(14.44)	(14.49)	(14.55)	(14.56)
oct	55.95***	51.31***	50.40***	49.45***	49.31***	48.27***	48.43***	45.47***
	(14.21)	(14.12)	(14.14)	(14.17)	(14.20)	(14.25)	(14.30)	(14.32)
nov	51.51***	46.41***	46.49***	46.29***	46.78***	45.88***	46.07***	43.39***
	(14.36)	(14.28)	(14.30)	(14.34)	(14.38)	(14.43)	(14.48)	(14.49)
free1	161.5***	164.0***	164.0***	165.4***	165.6***	164.1***	164.2***	161.3***
	(57.00)	(56.48)	(56.40)	(56.38)	(56.41)	(56.48)	(56.56)	(56.46)
free2	277.4***	278.4***	277.9***	275.2***	274.6***	273.8***	273.7***	273.3***
	(38.34)	(37.99)	(37.94)	(37.95)	(38.83)	(38.87)	(38.91)	(38.83)
free3	190.0***	190.1***	189.6***	187.9***	189.3***	189.2***	189.2***	191.9***
	(52.10)	(51.64)	(51.57)	(51.55)	(57.17)	(57.20)	(57.26)	(57.16)
free4	251.5***	252.4***	256.5***	255.3***	257.2***	257.7***	257.7***	257.1***
	(50.22)	(49.77)	(49.75)	(49.72)	(49.82)	(49.86)	(49.90)	(49.80)
Constant	1,413***	1,408***	1,423***	1,436***	1,441***	1,444***	1,445***	1,455***
	(60.11)	(59.97)	(60.27)	(60.69)	(61.69)	(61.88)	(62.24)	(62.46)
Observations	1,064	1,063	1,062	1,061	1,060	1,059	1,058	1,057
R-squared	0.773	0.777	0.778	0.778	0.778	0.778	0.778	0.779
Adj. R-squared	0.766	0.770	0.771	0.771	0.771	0.771	0.770	0.771

Table D: Result of regressing using the social media index while including lead variables, as discussed in section 8.4.3. Beta coefficients are marked differently depending on their level of significance. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Appendix E–Amount of patients' behavior explained by indices

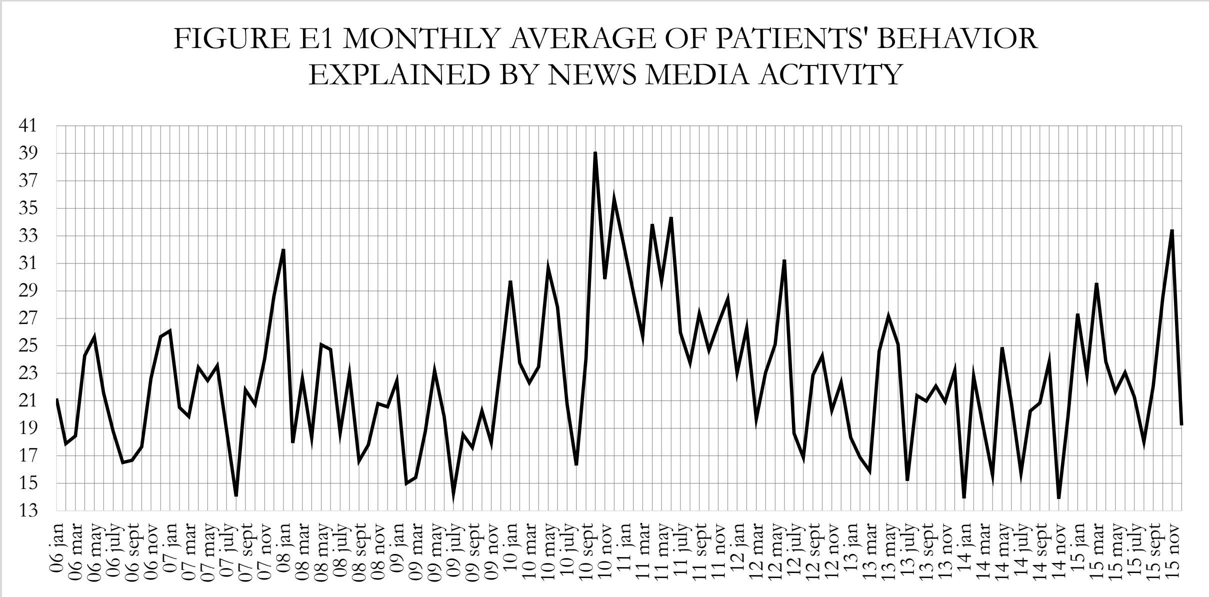

Figure E1: Given a beta coefficient of 0.301 in the regression and an average of 75 points of the index difference day to day, 23 patients' behavior can be explained by societal apprehension as measured by news media activity. The figure shows how the monthly average of this number fluctuates over time.

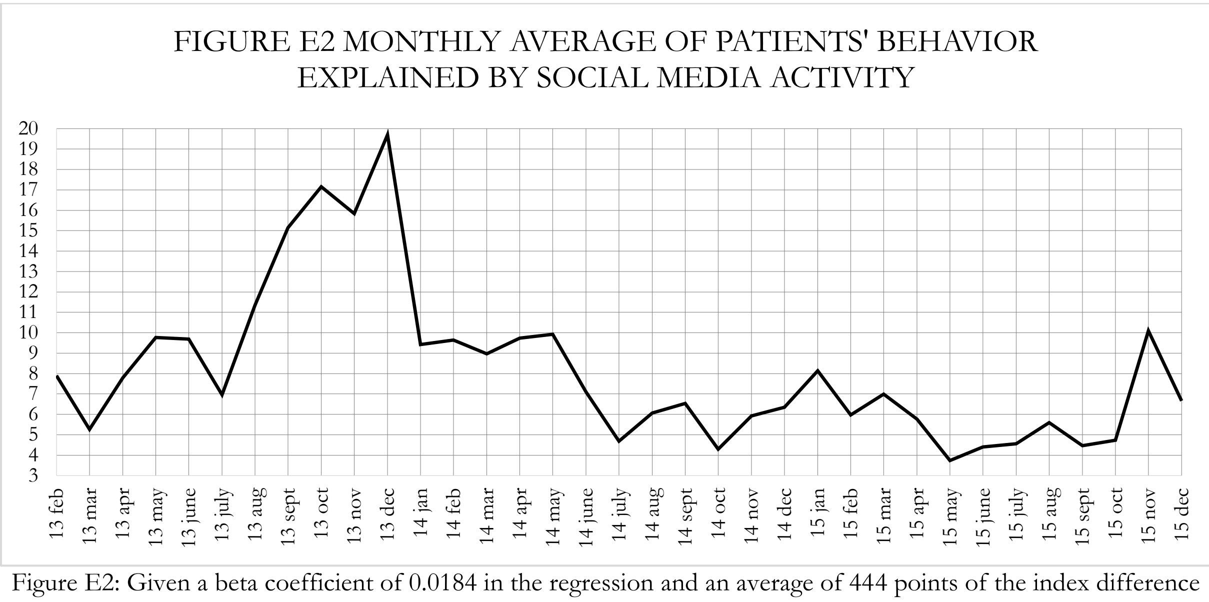

Figure E2: Given a beta coefficient of 0.0184 in the regression and an average of 444 points of the index difference day to day, 8 patients' behavior can be explained by societal apprehension as measured by social media activity. The figure shows how the monthly average of this number fluctuates over time.