**Big Data CA2**

## Goal-setting and Application Understanding

Social disorganization theory posits that the structural composition of neighbourhoods has an influence on the level of crime in that area (Stark, 1987). This theory suggests that the physical make-up of the area influences people’s behaviour, with features such as higher population density, poverty, and dilapidation leading to higher crime rates. Therefore, crime rates in an area are believed to be influenced not just by human factors such as the age and gender of the residents living there, but can also be a product of the physical environment (Shaw & McKay, 1942).

In support of this theory, there has been a large number of studies connecting the spatial distribution of alcohol outlets with the level of crime in an area (Franklin et al., 2010; Gorman et al., 2001; Gorman et al., 1998; Livingston, 2008; Yu et al., 2009). There is a positive association between alcohol consumption and the perpetration of crime, mediated somewhat by the behavioural disinhibition caused by drinking alcohol (Sontate et al., 2021). Estimates in America suggest that alcohol and drugs are implicated in up to 80% of crimes leading to incarceration, with common offenses including domestic violence, drink driving, property crimes and public-order infractions (National Institute on Drug Abuse et al., 2020). As alcohol is the most widely used drug causing disinhibition, it plays a large role in crime, even more so than most illegal drugs (Harris & Lehman, 2022). Previous research has indicated that the odds of becoming involved in crime increase significantly just after individuals reach the legal drinking age (Carpenter & Dobkin, 2015). Exploring this link between alcohol and crime in an Irish context is valuable for a number of reasons.

Ireland has one of the highest per-capita alcohol consumption rates in the European Union (Manthey et al., 2019), and a culture that promotes frequent drinking occasions and is broadly accepting of excessive ‘binge’ drinking (Ramstedt & Hope, 2003; Taber et al., 2019). Given the high volume pattern of consumption and binge drinking, the association between alcohol outlet density and crime rates may play out in a unique way in the Irish context. This analysis may provide valuable insight into crime and alcohol consumption trends across Ireland.

*Our Question:*

Is there an association between the rates of particular types of crime and the availability of alcohol?

Is this consistent across the country?

*Our Hypothesis:*

We predict to find a positive association between alcohol outlet density and overall crime rates. We expect this to be consistent across the country.

## Data Selection and Integration

### The Data

**1) Crime Data – roughly 135,000 rows of data.**

(Data from https://ws.cso.ie/public/api.restful/PxStat.Data.Cube\_API.ReadDataset/CJA07/XLSX/2007/en)

*The dataset used was sourced from the Central Statistics Office of Ireland and includes the number of crime offences under different categories. Columns include the type of crime, the total number of that crime in a given year, and the Garda station that crime was reported to. This dataset includes data from 2003 up to 2022. It is important to note that this crime data is considered “Under Reservation” due to concerns with the quality of the recorded crime data. Further revisions to the data can be expected as the underlying data quality deficiencies are resolved.*

**2) Alcohol Licence Data – just shy of 14,000 permits.**

(Data from <https://www.revenue.ie/en/corporate/documents/statistics/excise/liquor-licences.csv>)

 This dataset related to the register of liquor licences in Ireland in 2021/2022. Each row of the dataset related to a specific Business or Purveyor, and included a full address for the business. We extracted the ‘county’ from this address and used this for our analysis – matching them with the Garda Divisions category from the Crime dataset. Unfortunately we were unable to validate if all of the licences valid for 2021/22 were also valid in the years prior, which may have an impact on our analysis as we are examining crime data from multiple years.

**3) Population Data**

(Data from https://ws.cso.ie/public/api.restful/PxStat.Data.Cube\_API.ReadDataset/FP001/XLSX/2007/en)

This dataset was taken from the 2022 Census in Ireland and breaks down population by each county.

### The Method

For the purposes of this analysis, the data was aggregated into the Garda Division categories, with alcohol outlet density and crime rates for each division normalised for the population of that division. We then looked at the relationships between these variables to determine if alcohol outlet density may be an influencing factor in crime rates in Ireland.

There are a number of different methods used in research for measuring spatial alcohol density. Counts are the simplest measure, which capture the number of alcohol outlets available in a given area by summing the number of outlets in that region (Handy & Niemeier, 1997). Other measures, such as mean distance measures use descriptive statistics to summarize the distance from a given location to an outlet (Trangenstein et al., 2019). We used counts for our analysis.

For our analysis we categorised the alcohol and crime data by Garda divisions, then controlled for the populations of those divisions. We then looked to see if there was a significant positive correlation between alcohol outlet density and crime rates, which may indicate that increased availability of alcohol could lead to heightened crime rates.

## Data Cleaning & Pre-processing

See Google Colab Notebook.

## Data Transformation

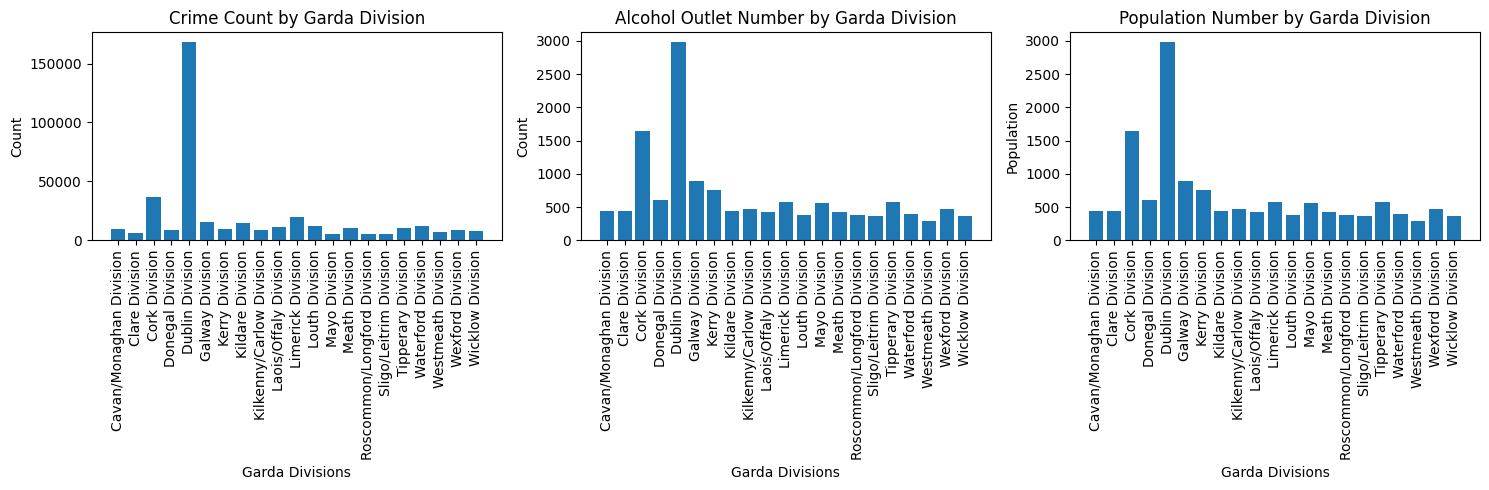
See Google Colab Notebook.

## Data Mining

See Google Colab Notebook.

## Pattern Evaluation / Interpretation

*1) Crime Rates:* Over the two year period (2021 & 2022) the highest crime rates were in the Dublin Division (168,135), followed by Cork (36,300), Limerick (19,453) and Galway (15,512). This is to be expected as these are some of the most populous Garda Divisions, and crime rates are typically higher in urban areas (Ladbrook, 1988; Onyeneke & Karam, 2022).

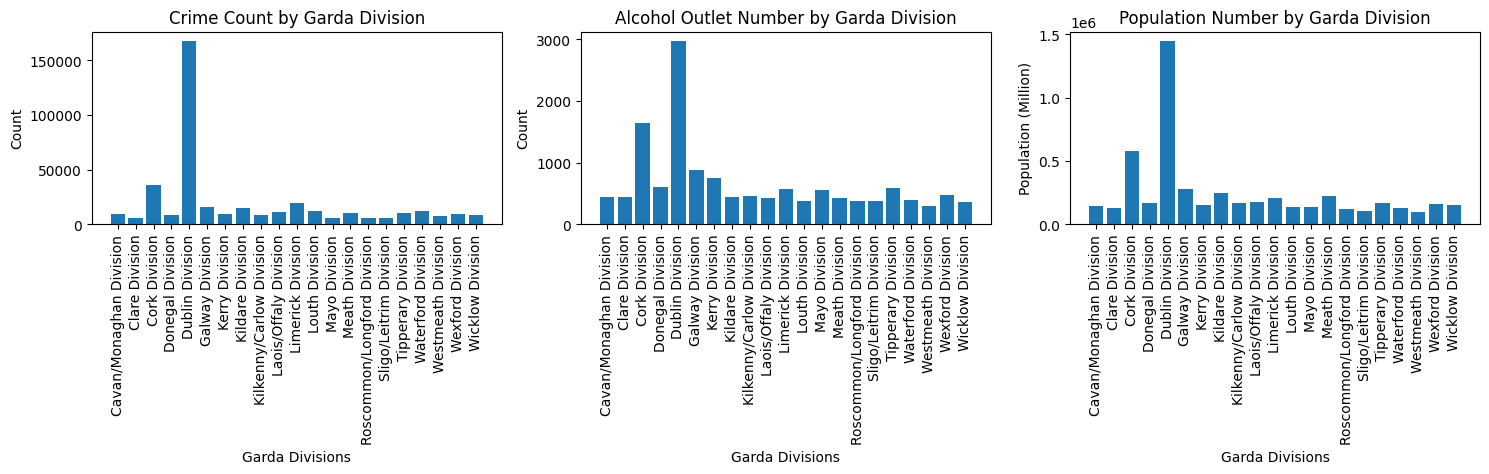


The crime types were categorised into a number of different types:

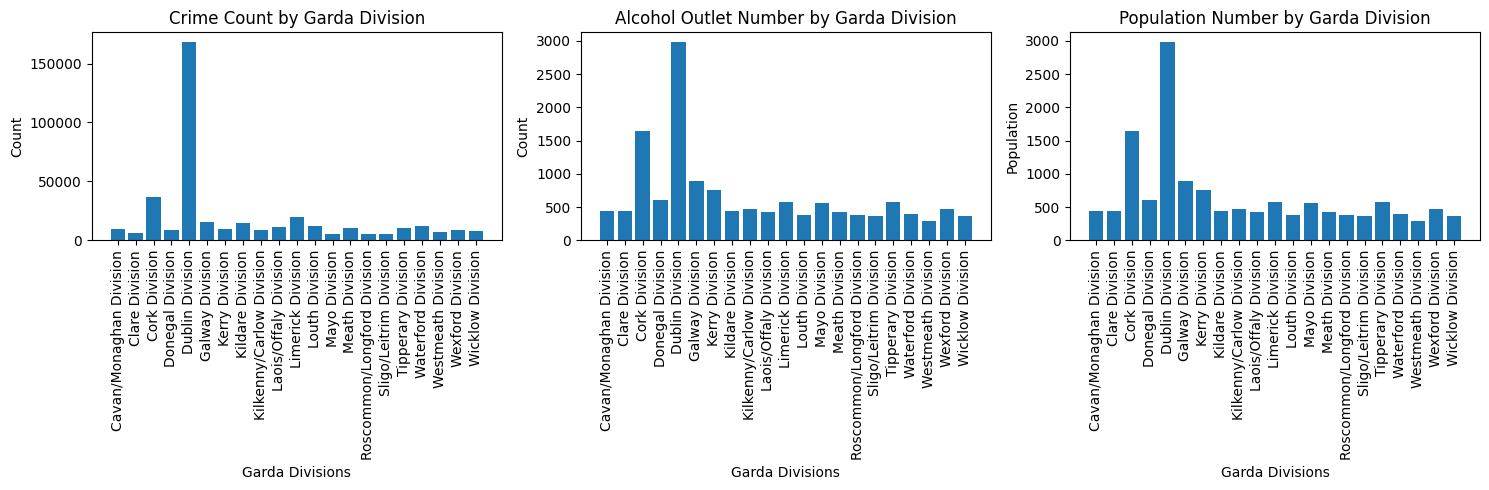


Theft and related offences were by far the most common crimes committed over 2021 and 2022, followed by Public Order and Social Code offences, and Assault offences. Evidence indicates that thefts, assaults, and public disorder crimes are all promoted by alcohol availability and excessive alcohol consumption (Harris & Lehman, 2022).

*2) Population Data:* The highest populations are seen in the urban areas such as the Dublin Division (1,450,701), followed by Cork (581,231), Galway (276,451), Kildare (246,977), Meath (220,296) and Limerick (205,444).

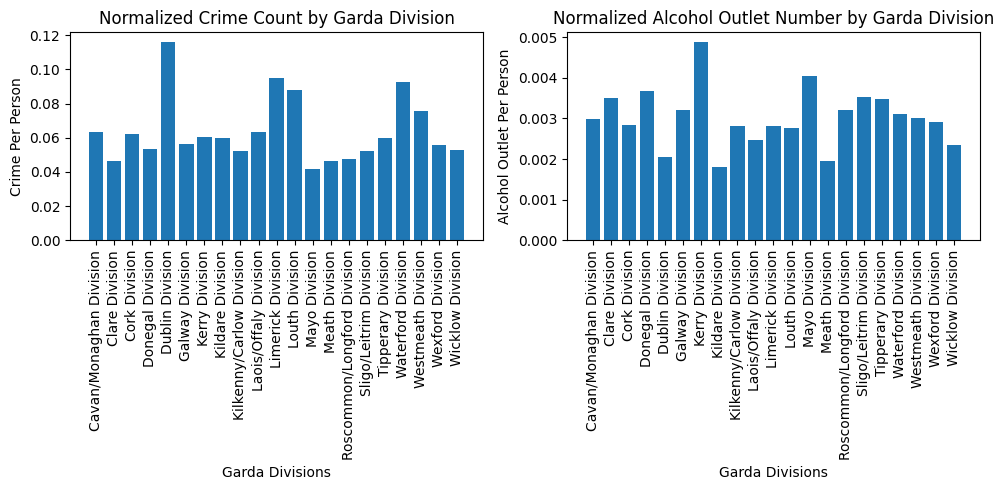


*3) Alcohol Outlet Data:* The highest number of alcohol outlets were seen in the Dublin Division (2,980), followed by Cork (1,648), Galway (886), Kerry (757) and Donegal (612). It is interesting to note that Meath (14th most outlets) and Kildare (11th most outlets) have high populations but not high alcohol outlet density. The fact that these are typically commuter counties with a large population who may work and socialise in Dublin may be a factor in why they do not have a high number of alcohol outlets. Geographic considerations like this are important to keep in mind when exploring the data.

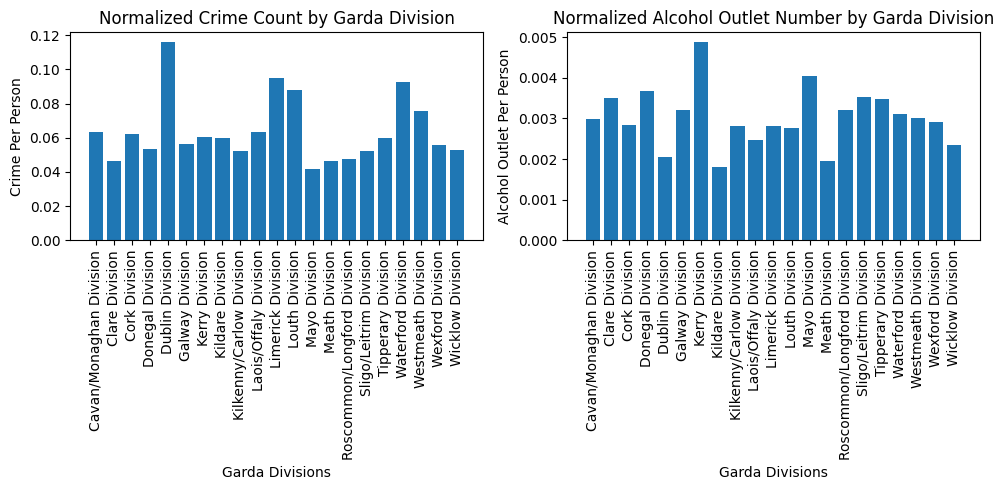


As expected, crime rates and alcohol outlet densities are typically linked to the population, so normalising the counts to account for population may provide a more accurate picture.

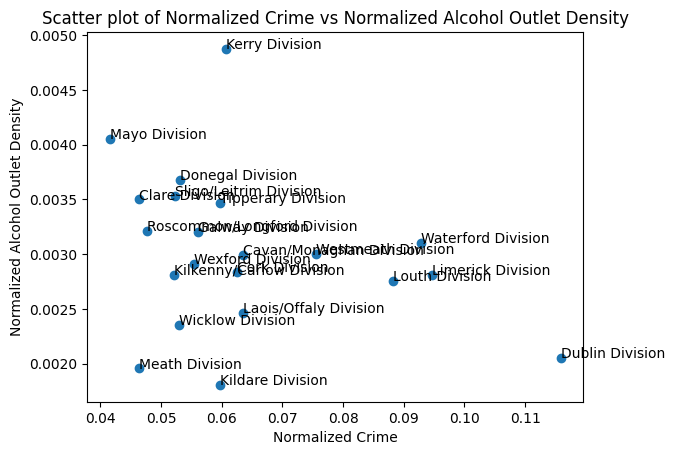
*4) Normalised Crime Counts:* When accounting for the population of the division, the top crime counties changes. Dublin still tops the list with 0.12 crimes per person (per 2 years), followed by Limerick (0.09), Waterford (0.09), Louth (0.09), and Westmeath (0.08).



*5) Normalized Alcohol Outlet Density:* When accounting for the population of the division, Kerry has the highest alcohol outlet density with 0.0049 alcohol outlets per person, followed by Mayo (0.0041), Donegal (0.0037), and the Sligo/Leitrim division (0.0035).



A scatterplot is a useful way to explore the relationship between crime rates and alcohol outlet density in the divisions.



As we can see in the scatterplot, there appears to be a weakly negative correlation between alcohol outlet density and crime rates in the Garda Divisions. Dublin and Kerry appear to be outliers which were removed for further analysis. However, the data still did not meet the assumptions of normality and homoscedasticity that are required for a correlation co-efficient so the validity of the result is questionable. The correlation co-efficient was -0.183, indicating a weakly negative correlation between the number of alcohol outlets in a division and the crime rate within that division i.e. as the alcohol outlet density increases, the crime rate tends to decrease slightly. It is important to note that correlation does not imply causality. This result does go against our initial hypothesis that suggested that more availability of alcohol would lead to higher crime rates. Due to the violations of assumptions in the data we decided not to run a linear regression. It is important to note that this analysis looks at ALL crime - different crime types may be affected differentially by alcohol availability, for example, drunk and disorderly offences should be more impacted than fraud.

### Knowledge Discovery & Use

In conclusion, our analysis did not support the hypothesis that alcohol outlet density leads to increased crime rates in Ireland. However, a number of limitations exist with our analysis so further exploration of the topic is warranted.

Besides the limitations mentioned when discussing the data quality, distribution etc, additional limitations to bear in mind with the analysis are that the alcohol licence data simply reveal where the outlets are located. They provide no information about how much the outlet is actually used or what volume of alcohol is sold at the premises. Quantity does not necessarily imply usage. Furthermore, another pitfall of this type of analysis is that we treat each Garda Division as a separate entity. The problem with this is that there may be bars and shops in neighbouring Garda Divisions that are used by the residents of a different division.

For the purposes of this project, given issues with the alcohol outlet data e.g. they were not always detailed addresses, we used the count measure for each crime division. There is a problem with this type of analysis as it implicitly assumes that the points (outlets) are uniformly distributed across the area, and makes no allowances for the fact that outlets may be concentrated in one area. Given a cluster of alcohol outlets, the count method may indicate a high availability of alcohol to the population within that division, even if the majority of the population do not have easy access to this cluster.

In future, it would be useful to categorize the crime rates and the number of alcohol outlets in a more granular way, such as by neighbourhood or community, instead of at the large Garda Division level. This could provide useful information as if problematic areas could be identified, preventative steps could be taken to reduce crime through increased alcohol control, such as limiting sale times. An example where prohibitory steps such as this have worked is in Sáu Paola, Brazil, where laws mandating earlier closing times for bars resulted in a significant reduction in violent crime (Biderman et al., 2010). Additionally, exploring which crime types are most correlated with alcohol outlet density may provide useful information to be used by policymakers to try and prevent alcohol-related crime.

Furthermore, looking in more detail at the individual alcohol outlet may provide a more nuanced analysis, for example a cinema that has a licence to sell wine would most likely have a very different effect on crime than a raucous nightclub. Categorising the alcohol outlets by the type of premise they are may provide more useful insights. It would also be interesting to look at the demographics committing crimes, as the tendency to commit crimes after drinking may not be spread evenly across the population, with research suggesting the alcohol/crime link is more pronounced for young males, who are more likely to drink heavily and more likely to commit crime in general (Harris & Lehman, 2022). Therefore, while this analysis did not find a positive correlation between crime rates and alcohol outlet density, there is still much more research that can be done to fully understand the relationship between these two variables in Ireland.

**References**

Biderman, C., De Mello, J. M., & Schneider, A. (2010). Dry laws and homicides: evidence from the São Paulo metropolitan area. *The economic journal*, *120*(543), 157-182.

Carpenter, C., & Dobkin, C. (2015). The minimum legal drinking age and crime. *Review of economics and statistics*, *97*(2), 521-524.

Franklin, F. A. I., LaVeist, T. A., Webster, D. W., & Pan, W. K. (2010). Alcohol outlets and violent crime in Washington DC. *Western Journal of Emergency Medicine*, *11*(3), 283.

Gorman, D. M., Speer, P. W., Gruenewald, P. J., & Labouvie, E. W. (2001). Spatial dynamics of alcohol availability, neighborhood structure and violent crime. *Journal of studies on alcohol*, *62*(5), 628-636.

Gorman, D. M., Speer, P. W., Labouvie, E. W., & Subaiya, A. P. (1998). Risk of assaultive violence and alcohol availability in New Jersey. *American journal of public health*, *88*(1), 97-100. <https://doi.org/10.2105/ajph.88.1.97>

Handy, S. L., & Niemeier, D. A. (1997). Measuring accessibility: an exploration of issues and alternatives. *Environment and planning A*, *29*(7), 1175-1194.

Harris, C., & Lehman, C. F. (2022). Fixing Drinking Problems: Evidence and Strategies for Alcohol Control as Crime Control. *Manhattan Institute, May*, *25*.

Ladbrook, D. A. (1988). Why are crime rates higher in urban than in rural areas?—Evidence from Japan. *Australian & New Zealand Journal of Criminology*, *21*(2), 81-103.

Livingston, M. (2008). Alcohol outlet density and assault: a spatial analysis. *Addiction*, *103*(4), 619-628.

Manthey, J., Shield, K. D., Rylett, M., Hasan, O. S., Probst, C., & Rehm, J. (2019). Global alcohol exposure between 1990 and 2017 and forecasts until 2030: a modelling study. *The Lancet*, *393*(10190), 2493-2502.

National Institute on Drug Abuse , N. i. o. H., Health, U. D. o., & Services, H. (2020). Criminal justice DrugFacts. In.

Onyeneke, C. C., & Karam, A. H. (2022). An exploratory study of crime: examining lived experiences of crime through socioeconomic, demographic, and physical characteristics. *Urban Science*, *6*(3), 43.

Ramstedt, M., & Hope, A. (2003). The Irish drinking culture. *Dublin: Health Promotion Unit, Department of Health and Children*.

Shaw, C. R., & McKay, H. D. (1942). Juvenile delinquency and urban areas.

Sontate, K. V., Rahim Kamaluddin, M., Naina Mohamed, I., Mohamed, R. M. P., Shaikh, M. F., Kamal, H., & Kumar, J. (2021). Alcohol, aggression, and violence: from public health to neuroscience. *Frontiers in Psychology*, *12*, 699726.

Stark, R. (1987). DEVIANT PLACES: A THEORY OF THE ECOLOGY OF CRIME. *Criminology*, *25*(4), 893-910. <https://doi.org/10.1111/j.1745-9125.1987.tb00824.x>

Taber, M., Stokes, A., Gallagher, N., & McClafferty, C. (2019). A report prepared by Behaviour & Attitudes for Drinkaware in Ireland.

Trangenstein, P. J., Curriero, F. C., Jennings, J. M., Webster, D., Latkin, C., Eck, R. H., & Jernigan, D. H. (2019). Methods for Evaluating the Association Between Alcohol Outlet Density and Violent Crime. *Alcohol Clin Exp Res*, *43*(8), 1714-1726. <https://doi.org/10.1111/acer.14119>

Yu, Q., Li, B., & Scribner, R. A. (2009). Hierarchical additive modeling of nonlinear association with spatial correlations—An application to relate alcohol outlet density and neighborhood assault rates. *Statistics in Medicine*, *28*(14), 1896-1912.