**Team:** The Wilderpeople

**Project:** Gravediggers

**Data analysis component:**

**Aim:** To calculate an approximate number of years of life lost as a result of exposure to risk-taking behaviour; based on the assumption that choices you make today will be similar to the choices that you make in the future, and that these choices have a cumulative impact on your life expectancy.

**Method and assumptions:** For the purposes of prototyping, we selected 5 risk-taking behaviours (alcohol use, tobacco smoking, unsafe sex, physical activity and fruit/vegetable intake) that are generally publically well-known and supported by high profile government-sponsored advertising campaigns. These behaviours are relevant to our users’ everyday lives and had high quality data available. Future development plans for the game would incorporate more risk factors (including injury risk) from more data sources.

The main data source for the calculations in Gravediggers is the [New Zealand Burden of Disease Study](http://www.health.govt.nz/nz-health-statistics/health-statistics-and-data-sets/new-zealand-burden-diseases-injuries-and-risk-factors-study-2006-2016) (NZBD). From the datasets released, we extracted the relevant risk-taking categories by age and gender, where our key outcome measure was years of life lost (YLL). Specific methodology for the calculation of YLL can be found in the method information available for the NZBD.

We made the assumption that where the risk taking behaviour was not engaged, the YLL value was 0, and that not everyone in each age/gender category would be engaging in the risk-taking behaviour (and therefore wouldn’t have contributed to the YLL related to that risk). Therefore the population in the NZBD data was different to the true population at risk (PAR). To generate a PAR for each age/gender category by risk, we matched the age and gender categories to the [New Zealand Health Survey](http://www.health.govt.nz/publication/portrait-health-key-results-2006-07-new-zealand-health-survey) and its supplements ([Alcohol](http://www.health.govt.nz/nz-health-statistics/national-collections-and-surveys/surveys/current-recent-surveys/alcohol-and-drug-use-survey) and [Tobacco](http://www.health.govt.nz/system/files/documents/publications/nz-tobacco-use-survey-2006-v2.doc) use), and the [Durex Safe Survey](http://www.data360.org/pdf/20070416064139.Global%20Sex%20Survey.pdf), to determine a relevant population size for each risk taking category, from which the YLL were lost. We used prevalence information from these surveys and applied this rate to the population size from the NZBD. We matched populations for dates as close as possible to the NZBD source data (2006/07) under the assumption that the PAR was measured at the start of the time period where YLL were accumulated (and therefore PAR members were eligible to contribute to YLL), and used NZBD source data where possible.

Where age-group categories did not completely match, aggregation of age-groups for the PAR was carried out (e.g. age groups 15-24 and 25-34 had to be aggregated into 0-34). The main grouping issue was for the younger age groups. In general, prevalence for risk taking behaviours was not statistically significantly different between these groups, so a mean prevalence was calculated for each age/gender category that needed to be aggregated. We did not weight for population size in these categories due to time constraints but would complete this in future development.

From these assumptions and the previous calculations, we produced a table relating risk-taking behaviour to age/gender groupings, with YLL and PAR for each category of age, gender and risk-taking behaviour.

Because YLL is a cumulative measure over the whole PAR, we then made the assumption that there could be an average number of years of life lost per individual within the PAR. This means that we assume that each member of the age/gender category who is exposed to the risk taking behaviour is equally likely to contribute to YLL. Future iterations of Gravediggers could include further category inputs, for example, ethnicity or co-morbidities, to give a more accurate estimation of the effect of risk-taking behaviour.

The choices made earlier in life also impact the likelihood of death or disability later in life (e.g. cancer as a result of smoking doesn’t present until later in life). We made the assumption that YLL/PAR could be added cumulatively for age groups after the user-selected age-group. For example, if the user selects a 0-34 age category, YLL for this category and all future ages would be added together, to estimate the effect of continuing to make decisions similar to that chosen in the game (CumulYLL). This is because the impact of risk-taking behaviour is often not experienced until later years and needs to be taken into account in our users’ decision making. This method has been designed by the Wilderpeople’s Analysts, and has not been tested for robustness, however we believe it to be relatively sound conceptually, for the purpose of estimating and displaying to users to impacts of their risk-taking behaviours.

Each risk-raking behaviour generates a different CumulYLL. The CumulYLL value for each age/gender/risk grouping is selected based on the user’s choices within the game and stored for presentation in the end-of-game dashboard. We assumed that the measure for each risk-taking behaviour is independent enough from other risk-taking behaviours to enable addition of CumulYLL for each risk-taking behaviour to come up with an estimate for a total life-lost measure for the impact of choices made within the game. Future development would require determining whether this is a sound assumption.

Facts within the game are presented for educational purposes and are generally extracted from pre-published information. We did not carry out any analysis on this data, but believe its reuse increases the visibility of government data.

We also began to collect information on prevalence data for diseases related to our key risk-taking behaviours. Our intention was to relate the choices of the users within the game to the different disease rates. We would provide a personalised word-cloud with the sizes of each word in the cloud based on the likely prevalence (or incidence) rates (based on data available) related to the risk-taking behaviours chosen in the game. We were not able to complete this work but would given sufficient development time, resources and data sources.

**Sources:** Sources that we used for analysis and “fact” development are:

<http://www.sfc.org.nz/infohealtheffects.php>

<https://www.health.govt.nz/system/files/documents/publications/health-loss-in-new-zealand-final-v2.pdf>

<http://www.health.govt.nz/system/files/documents/publications/ways-and-means-report-on-methodology-from-nzbds-aug13-v2.pdf>

<http://www.health.govt.nz/publication/new-zealand-burden-diseases-statistical-annexe>

<http://www.health.govt.nz/nz-health-statistics/health-statistics-and-data-sets/new-zealand-burden-diseases-injuries-and-risk-factors-study-2006-2016>

<http://www.health.govt.nz/publication/portrait-health-online-data-tables-2006-07-new-zealand-health-survey-results>

<http://www.health.govt.nz/nz-health-statistics/national-collections-and-surveys/surveys/current-recent-surveys/alcohol-and-drug-use-survey>

<http://www.health.govt.nz/system/files/documents/publications/nz-tobacco-use-survey-2006-v2.doc>

<http://www.data360.org/pdf/20070416064139.Global%20Sex%20Survey.pdf>

<http://www.hpa.org.nz/sites/default/files/Attributable%20fractions%20Final.pdf>

<http://smokefree.org.nz/smoking-its-effects/cost-of-smoking/cost-of-smoking-calculator>

<http://www.ash.org.nz/wp-content/uploads/2013/01/Factsheets/08_Reproductive_health_ASH_NZ_factsheet.pdf>

All analysis and data is made available in a separate data file, including logic for how we expect the app interface to connect with a database. This connection was not fully developed due to time and development constraints over GovHack, but we believe it to be feasible given appropriate resources.