# A simple method for correcting for the Will Rogers phenomenon with biometrical applications

**Mark Stander**[*1] and **Julian Stander** [2]

[1] BAE Systems
6 Carlton Gardens
LONDON
SW1Y 5AD
United Kingdom
Email: mark.stander@baesystems.com

[2] School of Engineering, Computing and Mathematics
University of Plymouth
Drake Circus
PLYMOUTH
PL4 8AA
United Kingdom
Email: J.Stander@plymouth.ac.uk

In its basic form the Will Rogers phenomenon takes place when an increase in the average value of each of two sets is achieved by moving an element from one set to another. This leads to the conclusion that there has been an improvement, when in fact essentially nothing has changed. Extended versions of this phenomenon can occur in epidemiological studies, rendering their results unreliable. After describing epidemiological and clinical studies that have been affected by the Will Rogers phenomenon, this paper presents a simple method to correct for it. The method involves introducing a transition matrix between the two sets and taking probability weighted expectations. Two real-world biometrical examples, based on migration economics and breast cancer epidemiology, are given and improvements against a naïve analysis are demonstrated. In the cancer epidemiology example we take account of estimation uncertainty. We also discuss briefly some limitations associated with our method.

---

*Corresponding author: e-mail: mark.stander@baesystems.com

# 1    Introduction

The Will Rogers phenomenon takes its name from the following quotation, attributed to the Oklahoma actor and personality Will Rogers:

> When the Okies left Oklahoma and moved to California, they raised the average intelligence level in both states.
>
> Will Rogers

In this tongue in cheek quotation, Will Rogers is effectively suggesting that the Okies family are of below average intelligence in Oklahoma, but of above average intelligence in California. So, when they move from Oklahoma to California, Oklahoma loses a below average intelligence family and hence its average intelligence increases, while California gains an above average intelligence family and so its average intelligence also increases. If the average state intelligence were reported before and after the Okies' move, there is an increased intelligence level in both states, possibly suggesting that over this time period intelligence has increased overall. However, the reality is that there has been no increase in overall intelligence. What is reported is essentially an artefact of the Okies' move.

The example that we have just discussed is, of course, not a serious one. However, there is a range of serious biometrical situations in which the Will Rogers phenomenon is observed including those concerning migration economics and cancer epidemiology. We motivate our work in Section 1.1 by outlining examples of where the Will Rogers phenomenon has had an effect on the results of epidemiological studies. In Sections 1.2 and 1.3 we present some mathematical notation and a simple example. We discuss a general method to correct for the Will Rogers phenomenon in Section 2. We then describe two worked examples of our method based on the real-world scenarios of migration economics and medicine in Section 3. In our medical example we also show how to take into account the uncertainty associated with estimated survival rates. We conclude with a brief discussion in Section 4. We separately provide R code to reproduce some of the key calculations. We now motivate our work by showing that the Will Rogers phenomenon is actually a problem that needs to be corrected in a range of epidemiological studies.

## 1.1    Motivation for the need to correct for the Will Rogers phenomenon in epidemiological studies

In medicine, the Will Rogers phenomenon is well recognised and refers to the apparent 'improved' survival of patients with cancer (Tan et al., 2015; Albertsen et al., 2005; Feinstein et al., 1985; Gofrit et al., 2008; Qi and Moul, 2017) or other diseases including multiple sclerosis (Sormani, 2009) due to reclassifying them into different prognostic groups that recognise more subtle disease manifestations, or to using different diagnostic modalities that allow the disease to be identified earlier. The Will Rogers phenomenon results in a 'zero-time shift' and improved prognosis without affecting actual survival (Segen, 2012). The 'zero-time' of a disease in prognosis studies is a well-defined point in time, such as the onset of symptoms or the beginning of treatment, which is set as a starting point. Changes in zero-time therefore can result in different prognosis, even if every patient had the same disease course (Fletcher and Fletcher, 2005). The Will Rogers phenomenon in oncology is also known as 'stage migration'. Stage migration is due to different methods of cancer staging which can artefactually alter some or all stage specific survival statistics by shifting patients with a marginal prognosis out of a better into a worse prognosis group (Blesch, 2005). It should be noted that staging is often used as an element of assigning a treatment regime. Also, if there has been no improvement in treatment or change in population structure, then the overall survival rate usually remains unchanged.

More sensitive diagnostic procedure can also lead to stage migration. For example, the use of CAT scanning and MRIs has resulted in the diagnosis of metastatic disease that would have been undetectable

some years ago. The patients who "move" due to this improvement in diagnostic procedure typically have worse survival than those with truly localized disease but better survival than those patients with obvious metastatic disease. The effect of reclassifying patients from the localized to the metastatic group will therefore result in apparent improvements in patient survival in both groups even when there has been no overall change in survival (Dickman and Adami, 2006).

We will now discuss several types of epidemiological study in which the Will Rogers phenomenon is recognised as a source of considerable potential bias. In studies that involve retrospective comparisons of contemporary to historical result such as Gofrit et al. (2008) which considers urological oncology, the Will Rogers phenomenon can be the main source of bias. Hence, correcting for it is key, if we wish to make historical comparisons of disease. Longer-term outcome studies, across which diagnostic criteria might have changed, would also benefit from a method that can correct for the associated bias. A study of carcinoma of the cervix (Foley et al., 2013), which compared survival pre-1999 and post-1999, provides an example. This study found that its results were subject to bias due to the 'classic Will Rogers phenomenon'. Similarly, the Will Rogers phenomenon was suggested as providing bias that accounted for 'at least some of the improved survival observed in consecutive trials' aimed at assessing differences in non-small-cell lung cancer outcomes in a French department from 1982 to 1997 (Foeglé et al., 2005). Long-term observation studies can also be affected by similar sources of bias. For example, the Will Rogers phenomenon is highlighted as a key methodological issue connected with the measurement of long-term treatment effects in multiple sclerosis (Sormani and Bruzzi, 2015). In addition, the Will Rogers phenomenon is one of the three key problems that would need to be eliminated before historical controls can be used as an acceptable method of evaluation (Feinstein, 1979). We should therefore adjust for the effects of the Will Rogers phenomenon, because comparison against historical controls rather than placebo can be beneficial, when, for example, an investigator would not want to prevent patients from having access to allegedly superior treatments (Clark and Leaverton, 1994). Other benefits of historical controls include being able to reduce sample sizes (Quan et al., 2017) and allowing more resources to be devoted towards the novel therapy, while retaining accurate estimates of effect (Viele et al., 2014).

The final type of study where the Will Rogers phenomenon can provide bias is investigations in which there is a geographical or spatial component. For example, significant bias was acknowledged in a study that examined survival rates of non-small-cell lung cancer patients in difference geographic regions (Grosclaude et al., 1995). The bias was due to the fact that different geographies employed differing diagnostic criteria, so leading to an erroneous prediction of higher survival rates in certain areas. Correcting for the Will Rogers phenomenon in this case would allow for better comparison to be made between geographies. All these specific examples motivate the need for a general method for correcting for the Will Rogers phenomenon in epidemiological studies. We discuss such a method and its limitations in the rest of the paper.

### 1.2 Mathematical notation

In order to describe our method for correcting for the Will Rogers phenomenon, we need some straightforward mathematical notation. Let $A$ and $B$ be two disjoint sets of observations on people, for example, and let $P = A \cup B$ be the set of observations on the entire population. We observe $A$, $B$ and $P$ at two distinct times: $t = 0$, and $t = 1$. At $t = 0$, the sets $A$, $B$ and $P$ take the form

$$
\begin{aligned}
A_0 &= \{a_{0,1}, a_{0,2}, \ldots, a_{0,m_0}\} \\
B_0 &= \{b_{0,1}, b_{0,2}, \ldots, b_{0,n_0}\} \\
P_0 &= A_0 \cup B_0 = \{p_{0,1}, p_{0,2}, \ldots, p_{0,n_0+m_0}\},
\end{aligned}
$$

where we use the general notation $a$ and $b$ to denote observation values from sets $A$ and $B$ respectively, and $p$ for observation values from the entire population set $P$. We use the general notation $m$ and $n$ to indicate the number of observations in sets $A$ and $B$ respectively. Later at $t = 1$, after transitions have taken place,

these sets become

$$
\begin{aligned}
A_1 &= \{a_{1,1}, a_{1,2}, ..., a_{1,m_1}\} \\
B_1 &= \{b_{1,1}, b_{1,2}, ..., b_{1,n_1}\} \\
P_1 &= A_1 \cup B_1 = \{p_{1,1}, p_{1,2}, ..., p_{1,n_1+m_1}\},
\end{aligned}
$$

where $m_0 + n_0 = m_1 + n_1$. We use $p_{0,j}$, $j = 1, \ldots, n_0 + m_0$, and $p_{1,j}$, $j = 1, \ldots, n_1 + m_1$, for the elements of the unions $A_0 \cup B_0$ and $A_1 \cup B_1$ for ease of notation later.

Associated with these sets are the following arithmetic means or averages:

$$
\bar{A}_0 = \frac{1}{m_0} \sum_{j=1}^{m_0} a_{0,j} \qquad\qquad \bar{B}_0 = \frac{1}{n_0} \sum_{j=1}^{n_0} b_{0,j} \qquad\qquad \bar{P}_0 = \frac{1}{n_0 + m_0} \sum_{j=1}^{n_0+m_0} p_{0,j}
$$

$$
\bar{A}_1 = \frac{1}{m_1} \sum_{j=1}^{m_1} a_{1,j} \qquad\qquad \bar{B}_1 = \frac{1}{n_1} \sum_{j=1}^{n_1} b_{1,j} \qquad\qquad \bar{P}_1 = \frac{1}{n_1 + m_1} \sum_{j=1}^{n_1+m_1} p_{1,j}
$$

and differences between these arithmetic means from $t = 0$ to $t = 1$:

$$
\Delta\bar{A} = \bar{A}_1 - \bar{A}_0 \qquad\qquad \Delta\bar{B} = \bar{B}_1 - \bar{B}_0 \qquad\qquad \Delta\bar{P} = \bar{P}_1 - \bar{P}_0.
$$

### 1.3 A simple numerical example of the Will Rogers phenomenon based on a dummy eco-logical example

An example of when the Will Rogers phenomenon occurs is the following: at time $t = 1$, an element $b$ with the property that $b < \bar{B}_0$ and $b > \bar{A}_0$ is moved from set $B$ to set $A$. Since $b$ is below the arithmetic mean of set $B_0$, moving it out of this set will increase the mean of the remaining values so that $\bar{B}_1 > \bar{B}_0$ with the consequence that $\Delta\bar{B} > 0$. Similar, since $b$ is above the arithmetic mean of set $A_0$, moving it into this set will increase the mean of the expanded set so that $\bar{A}_1 < \bar{A}_0$, again with the consequence that $\Delta\bar{A} > 0$. The reclassification of an individual from one set to another leads to an increase in the means of both sets.

Similarly, if at time $t = 1$ an element $b$ with the property $b > \bar{B}_0$ and $b < \bar{A}_0$ is moved from set $B$ to set $A$, then a *reduction* rather than an increase in the means of both sets occurs. A reduction or an increase in the means of both sets may also be achieved by more complicated reclassification moves.

Misleading interpretations can result from changes in set membership such as those just described. Let us consider as a simple example, a study that involves measuring the biomass of birds on an island. The island in question has North and South populations of birds, denoted as $A$ and $B$ respectively, comprising individual birds with observed biomass, at $t = 0$, as follows:

$$
A_0 = \{1, 2, 3\} \qquad\qquad\qquad B_0 = \{4, 5, 6\},
$$

with arithmetic means $\bar{A}_0 = 2$ and $\bar{B}_0 = 5$ that represent the average biomass of a bird in North and South parts of the island.

At $t = 1$ the bird with biomass value 4 migrates from $B$ to $A$ yielding

$$
A_1 = \{1, 2, 3, 4\} \qquad\qquad\qquad B_1 = \{5, 6\},
$$

with arithmetic means $\bar{A}_1 = 2.5$ and $\bar{B}_0 = 5.5$, so that $\Delta\bar{A} = 2.5 - 2 = 0.5$ and $\Delta\bar{B} = 5.5 - 5 = 0.5$ are both positive. By this move, the weights of the birds has appeared to increased in both parts of the island. However, this apparent increase is artefactual as the biomass of the birds has not changed since $\Delta\bar{P} = \bar{P}_1 - \bar{P}_0 = 3.5 - 3.5 = 0$.

Transitions from one set to another such as the one just illustrated can have considerable impacts on statistical summaries. A method that allows us to correct for such artefactual changes will be of use in several biometrical fields, including migration economics and medicine.

It is important to note that changing the definitions of sets can have the same effect as moving elements from one set to another, where that change in definition means that elements move from one classification to another. In the example above, we can achieve the same Will Rogers phenomenon effect by changing the definition of the sets, rather than explicitly moving elements from one set to the other. To demonstrate this, let us think of the above sets $A_0$ and $B_0$ as being defined as containing values less than or greater than $3.5$ respectively at $t = 0$. At $t = 1$ these classifications are changed so that $A$ and $B$ are defined as containing values less than or greater than $4.5$ respectively. Changing the definitions of sets means that elements are classified to different sets, which is, in effect, the same as moving those elements from one set to another. In the following sections we outline a possible method for correcting for the Will Rogers phenomenon.

## 2    Setting up a method to correct for the Will Rogers phenomenon.

Our method for correcting for the Will Rogers phenomenon is based on the following transition matrix $M$ which contains the probabilities of moving from state $a_{0,j}, j = 1, \ldots, m_0$ and $b_{0,j}, j = 1, \ldots, n_0$ at $t = 0$ to $a_{1,j}, j = 1, \ldots, m_1$ and $b_{1,j}, j = 1, \ldots, n_1$ at $t = 1$:

$$
M = \begin{array}{c} a_{0,1} \\ a_{0,2} \\ \vdots \\ a_{0,m_0} \\ \hline b_{0,1} \\ b_{0,2} \\ \vdots \\ b_{0,n_0} \end{array}
\left[
\begin{array}{c|c}
\overset{a_{1,1},a_{1,2},\ldots,a_{1,m_1}}{\begin{array}{c} \mathrm{Pr}(a_{0,j} \to a_{1,k}) \\ j = 1,\ldots,m_0; k = 1,\ldots,m_1 \end{array}} & \overset{b_{1,1},b_{1,2},\ldots,b_{1,n_1}}{\begin{array}{c} \mathrm{Pr}(a_{0,j} \to b_{1,k}) \\ j = 1,\ldots,m_0; k = 1,\ldots,n_1 \end{array}} \\
\hline
\begin{array}{c} \mathrm{Pr}(b_{0,j} \to a_{1,k}) \\ j = 1,\ldots,n_0; k = 1,\ldots,m_1 \end{array} & \begin{array}{c} \mathrm{Pr}(b_{0,j} \to b_{1,k}) \\ j = 1,\ldots,n_0; k = 1,\ldots,n_1 \end{array}
\end{array}
\right] \tag{1}
$$

As the elements of the rows of any transition matrix $M$ are probabilities, they have to be positive or zero. As some form of transition has to take place, the elements of $M$ have to sum to 1 across rows. Within-set transitions are allowed.

The correction works by applying transition probabilities set up in a matrix. When an element in one set at time $t = 0$ has been moved to another set at time $t = 1$, this transition matrix allows us to correct for this movement by taking into accounts its probability. We can use this matrix to calculate summary statistics that take the form of probability-weighted expected values and that are used to correct for the Will Rogers phenomenon. We will present specific numerical examples of this in Section 3.

Based on the transition matrix $M$, we define the expected values of $a_{0,j}, j = 1, \ldots, m_0$, and $b_{0,j}$, $j = 1, \ldots, n_0$, after transition at $t = 1$ as:

$$
\begin{aligned}
\mathrm{E}\left[a_{0,j}\right] &= \sum_{k=1}^{m_1} a_{1,k} \mathrm{Pr}(a_{0,j} \to a_{1,k}) + \sum_{k=1}^{n_1} b_{1,k} \mathrm{Pr}(a_{0,j} \to b_{1,k}),\ j = 1, \ldots, m_0, \\
\mathrm{E}\left[b_{0,j}\right] &= \sum_{k=1}^{m_1} a_{1,k} \mathrm{Pr}(b_{0,j} \to a_{1,k}) + \sum_{j=1}^{n_1} b_{1,k} \mathrm{Pr}(b_{0,j} \to b_{1,k}),\ j = 1, \ldots, n_0.
\end{aligned}
$$

We can now define the difference between the expected value of each element after transition at time $t = 1$ and the the element itself:

$$
\begin{aligned}
\Delta a_{0,j} &= \mathrm{E}\left[a_{0,j}\right] - a_{0,j},\ j = 1, \ldots, m_0, \\
\Delta b_{0,j} &= \mathrm{E}\left[b_{0,j}\right] - b_{0,j},\ j = 1, \ldots, n_0.
\end{aligned}
$$

This leads to the definition of the following three quantities that measure the average change in the elements of $A_0$, $B_0$ and $P_0$ after transition at time $t = 1$:

$$\widehat{\Delta A_0} \;=\; \frac{1}{m_0} \sum_{j=1}^{m_0} \Delta a_{0,j},$$

$$\widehat{\Delta B_0} \;=\; \frac{1}{n_0} \sum_{j=1}^{n_0} \Delta b_{0,j},$$

$$\widehat{\Delta P_0} \;=\; \frac{1}{m_0 + n_0} \left( \sum_{j=1}^{m_0} \Delta a_{0,j} + \sum_{j=1}^{n_0} \Delta b_{0,j} \right).$$

In the case of the ecology example based on an island with two populations of birds, for which $A_0 = \{1, 2, 3\}$, $B_0 = \{4, 5, 6\}$, $A_1 = \{1, 2, 3, 4\}$ and $B_1 = \{5, 6\}$, we know from the scenario that the transition matrix $M$ takes the form

$$M = \begin{array}{c} \\ \\ \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \\ \left[ \begin{array}{cccc|cc} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array} \right]. \end{array}$$

It is now easy to find that $\mathrm{E}\,[a_{0,1}] = 1 \times 1 = 1$, $\mathrm{E}\,[a_{0,2}] = 2 \times 1 = 2$, $\mathrm{E}\,[a_{0,3}] = 3 \times 1 = 3$, $\mathrm{E}\,[b_{0,1}] = 4 \times 1 = 4$, $\mathrm{E}\,[b_{0,2}] = 5 \times 1 = 5$ and $\mathrm{E}\,[b_{0,3}] = 6 \times 1 = 6$, from which $\Delta a_{0,1} = 1 - 1 = 0$, $\Delta a_{0,2} = 2 - 2 = 0$, $\Delta a_{0,3} = 3 - 3 = 0$, $\Delta b_{0,1} = 4 - 4 = 0$, $\Delta b_{0,2} = 5 - 5 = 0$ and $\Delta b_{0,3} = 6 - 6 = 0$, leading to $\widehat{\Delta A_0} = 0$, $\widehat{\Delta B_0} = 0$ and $\widehat{\Delta P_0} = 0$. These quantities tell us that the average expected change in the elements of $A_0$, $B_0$ and $P_0$ due to the transition is zero. This reflects the underlying reality! Hence, $\widehat{\Delta A_0}$, $\widehat{\Delta B_0}$ and $\widehat{\Delta P_0}$ are useful measures that correct for the Will Rogers phenomenon.

# 3  Worked examples of our Will Rogers phenomenon correction method based on real-world scenarios

This section considers two worked examples based on real-world scenarios where the Will Rogers phenomenon has a potential impact on the inference made about the underlying value of two or more sets. We supply as supplementary material R (R Core Team, 2019) code to perform our main calculations.

## 3.1  Will Rogers phenomenon in migration economics

Economic studies use income *per capita* as a primary metric. A common task is to assess the economic impact of migration upon average income. The recent empirical literature about the effect of immigrants on the wages of natives has provided a mixed set of results (Borjas, 2017; Ottaviano and Peri, 2012). The Will Rogers phenomenon may produce artefactual result here, as we show in the following simple example in which migration leads to a *reduction*, rather than to an increase in set means.

Let us assume that we have a migration economy $A$ from which workers move to a more developed economy $B$. Let $A_0 = \{2, 2.5, 2.6\}$ be the incomes of the individuals in $A$ and $B_0 = \{3, 4.5, 6\}$ be the incomes of individuals in $B$ at $t = 0$, as measured in tens of thousands of dollars. A migration takes place at $t = 1$, after which $A_1 = \{2, 2.5\}$ and $B_1 = \{3, 4, 4.7, 5.5\}$, where we assume that the unit of income has the same value in real terms. We have $\bar{A}_0 = 2.367$, $\bar{B}_0 = 4.5$, $\bar{A}_1 = 2.25$ and $\bar{B}_1 = 4.3$, so that $\Delta\bar{A} = 2.25 - 2.367 = -0.117$ and $\Delta\bar{B} = 4.3 - 4.5 = -0.2$. These reductions are in line with the assertion of Felbermayr et al. (2010) that incomes are trivially suppressed by the movement of workers, if immigrants are on average poorer than natives. In this case, however, there has been an overall increase in average incomes from 3.433 at $t = 0$ to 3.6167 at $t = 1$. This means that presenting $\Delta\bar{A}$ and $\Delta\bar{B}$ would be misleading. As in this case there has been an increase in the overall mean, we can think of this as an extension of the Will Rogers phenomenon.

Let us assume that we do not know exactly how incomes change for the individuals involved, but that we have some idea from previous censuses about this. Let us also assume that it is more likely that the people with higher incomes migrate (Docquier et al., 2011). These assumptions allow us to construct a transition matrix such as the following, which models that we are blind to changes in incomes inside the two economies, after migration has been taken into account:

$$
M = \begin{array}{c} \\ \\ 2 \\ \\ 2.5 \\ \\ 2.6 \\ \\ \\ 3 \\ \\ 4.5 \\ \\ 6 \end{array}
\begin{array}{c} \begin{array}{cccccc} 2 & 2.5 & 3 & 4 & 4.7 & 5.5 \end{array} \\
\left[ \begin{array}{cc|cccc}
\frac{1}{2} & \frac{1}{3} & \frac{1}{6} & 0 & 0 & 0 \\[4pt]
\frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & 0 & 0 \\[4pt]
\frac{1}{6} & \frac{1}{3} & \frac{1}{2} & 0 & 0 & 0 \\[4pt]
\hline
0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\[4pt]
0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\[4pt]
0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3}
\end{array} \right]
\end{array}
$$

The assumption of equal transition probabilities between income states within an economy for non-migrating people can be relaxed provided the rows of $M$ contain positive elements that sum to 1.

It follows that $\Delta a_{0,1} = 2.33 - 2 = 0.33$, $\Delta a_{0,2} = 2.5 - 2.5 = 0$ and $\Delta a_{0,3} = 2.67 - 2.6 = 0.07$ so that $\widehat{\Delta A_0} = 0.4/3 = 0.13$, and $\Delta b_{0,1} = 4.73 - 3 = 1.73$, $\Delta b_{0,2} = 4.73 - 4.5 = 0.23$ and $\Delta b_{0,3} = 4.73 - 6 = -1.27$ so that $\widehat{\Delta B_0} = 0.7/3 = 0.23$. These positive $\widehat{\Delta A_0} = 0.13$ and $\widehat{\Delta B_0} = 0.23$ values are more in line

with our intuition about what may have happened in the economies than the negative $\Delta \bar{A} = -0.117$ and $\Delta \bar{B} = -0.2$.

### 3.2   Will Rogers phenomenon in cancer epidemiology

In this section, we use real data to show how a study can provide biased results due to the Will Rogers phenomenon. We also illustrate how a correction for a long-term intervention study may be found using the method that we have presented.

Cancer epidemiology is a complex field of study and survival analyses have various limitations, as examined in Dickman and Adami (2006). Despite these issues, we have chosen this example to illustrate the use of our method since, as discussed in Section 1.1, the Will Rogers phenomenon is particularly relevant to studies on cancer survival, if they analyse survival trends over time. In general, cancers are classified into different stages depending on the extent to which the tumour has progressed, and the definitions of these cancer stages can change over time. Our example is based on the AJCC staging system which uses methodology developed by the American Joint Committee on Cancer (AJCC, 2019).

Tan et al. (2015) consider a cohort of $3147$ breast cancer patients, whose cancers were classified to one of four stages. They found that changing the classification methodology from the AJCC5 (Fleming et al., 1997) to the AJCC6 (Greene et al., 2002) system led to a seemingly higher survival rate for Stage 2 and 3 patients, as shown in Table 1.

| Cancer Stage | Five-year survival rates | |
| --- | --- | --- |
| | AJCC5 (confidence interval) | AJCC6 |
| 1 | 96.9 (94.7, 98.6) | 96.9 |
| 2 | 82.9 (80.9, 84.9) | 86.1 |
| 3 | 50.6 (46.2, 54.9) | 59.0 |
| 4 | 14.4 (11.1, 18.2) | 12.6 |

**Table 1**   The five-year percentage survival rates according to the AJCC5 and the AJCC6 classification systems reported in Tan et al. (2015). Here we report the confidence intervals given in Tan et al. (2015) for the percentage of patients who survive according to the AJCC5 classification system.

An explanation for these changes in survival percentages can be found from the following matrix $R$ of transition counts:

$$
R \;=\; 
\begin{array}{r}
\\
\text{Stage 1}_{\text{AJCC5}} \\
\text{Stage 2}_{\text{AJCC5}} \\
\text{Stage 3}_{\text{AJCC5}} \\
\text{Stage 4}_{\text{AJCC5}}
\end{array}
\begin{array}{c}
\begin{array}{cccc}
\text{Stage 1}_{\text{AJCC6}} & \text{Stage 2}_{\text{AJCC6}} & \text{Stage 3}_{\text{AJCC6}} & \text{Stage 4}_{\text{AJCC6}}
\end{array}\\
\left[
\begin{array}{cccc}
706 & 0 & 0 & 0 \\
0 & 1230 & 323 & 0 \\
0 & 0 & 511 & 20 \\
0 & 0 & 20 & 337
\end{array}
\right]
\end{array},
$$

(2)

the $(i, j)$-element of which is the number of patients classified in Stage $i$ by the AJCC5 system and in Stage $j$ by AJCC6.

From the matrix $R$ of transition counts we see that no Stage 1 patient has been re-classified. This is because the criteria that define Stage 1 are the same in the AJCC5 and AJCC6 systems. However, 323 out of the 1553 AJCC5 Stage 2 patients were reclassified as Stage 3 by the AJCC6 system because these patients had four or more axillary lymph nodes. This means that the more severely affected Stage 2 patients have been moved out of that stage by the AJCC6 system, with the result that those patients who remain classified as Stage 2 are the less severely affected ones and hence show a higher survival percentage. Similar considerations may apply to Stage 3, which under the AJCC6 system gains the 323 more severely affected Stage 2 patients, loses 20 of the more severely effected 531 AJCC5 Stage 3 patients, but also gains 20 of the less severely affected 357 AJCC5 Stage 4 patients, who had supraclavicular lymph nodes as the only 'metastatic' site. The significant changes in survival seen in Table 1 are essentially due to changes in diagnostic criteria and are examples of a more general form of the Will Rogers phenomenon. It would be naïve to conclude that survival for patients originally assigned to Stage 2, for example, has increased.

To correct for this more general form of the Will Rogers phenomenon, an extension of the transition matrix defined in equation (1) is required. For example, the following more general transition matrix, based on the structure of the matrix given in (2), extends the number of sets considered from two ($A$ and $B$) to four ($A$, $B$, $C$ and $D$), by using a straightforward generalisation of notation:

$$M = \begin{array}{c|c|c|c|c} & a_{1,1},a_{1,2},...,a_{1,m_1} & b_{1,1},b_{1,2},...,b_{1,n_1} & c_{1,1},c_{1,2},...,c_{1,p_1} & d_{1,1},d_{1,2},...,d_{1,q_1} \\ \begin{array}{c} a_{0,1} \\ a_{0,2} \\ \vdots \\ a_{0,m_0} \end{array} & P(a_{0,j} \to a_{1,k}) & P(a_{0,j} \to b_{1,k}) & P(a_{0,j} \to c_{1,k}) & P(a_{0,j} \to d_{1,k}) \\ \hline \begin{array}{c} b_{0,1} \\ b_{0,2} \\ \vdots \\ b_{0,n_0} \end{array} & P(b_{0,j} \to a_{1,k}) & P(b_{0,j} \to b_{1,k}) & P(b_{0,j} \to c_{1,k}) & P(b_{0,j} \to d_{1,k}) \\ \hline \begin{array}{c} c_{0,1} \\ c_{0,2} \\ \vdots \\ c_{0,p_0} \end{array} & P(c_{0,j} \to a_{1,k}) & P(c_{0,j} \to b_{1,k}) & P(c_{0,j} \to c_{1,k}) & P(c_{0,j} \to d_{1,k}) \\ \hline \begin{array}{c} d_{0,1} \\ d_{0,2} \\ \vdots \\ d_{0,q_0} \end{array} & P(d_{0,j} \to a_{1,k}) & P(d_{0,j} \to b_{1,k}) & P(d_{0,j} \to c_{1,k}) & P(d_{0,j} \to d_{1,k}) \end{array}$$

$$(3)$$

If we now take sets $A$, $B$, $C$ and $D$ to correspond to Stage 1, 2, 3 and 4 cancers and if we let $t = 0$ and $t = 1$ correspond to the AJCC5 and AJCC6 classification systems, we can derive from the matrix $R$ of transition counts given in (2) a transition matrix $M$ of the form (3):

| $M =$ | Stage 1 AJCC6 | Stage 2 AJCC6 | Stage 3 AJCC6 | Stage 4 AJCC6 |
|---|---|---|---|---|
| Stage 1 AJCC5 | 1 | 0 | 0 | 0 |
| Stage 2 AJCC5 | 0 | 0.79 | 0.21 | 0 |
| Stage 3 AJCC5 | 0 | 0 | 0.96 | 0.04 |
| Stage 4 AJCC5 | 0 | 0 | 0.06 | 0.94 |

The above matrix was generated by taking each element of $R$ and dividing it by the sum of the elements in its respective row. For example, $0.79 = 1230/(1230 + 323) = 1230/1553$ and $0.21 = 323/1553$, to two decimal places. The matrix therefore represents the probability of an element being reclassified when the AJCC6 system is used instead of the AJCC5 system.

Let us now assume that we want to assess intervention outcomes for breast cancer over a long period of time using a retrospective study. However, we find that the diagnostic re-classification described above has taken place during the time period of the study, introducing Will Rogers phenomenon bias into the results. We must therefore correct for this bias.

We use the AJCC5 survival data given in Table 1 as a historical baseline, against which further comparison can be made. Let us say that, after an intervention, patients are assessed under the AJCC6 classification system as being in breast cancer Stages 1, 2, 3 and 4, and that the resulting four patient groups, which for simplicity are assumed to contain equal numbers of patients, are found to have the following five year survival percentages:

$A_1 = \{98\}$, $B_1 = \{85\}$, $C_1 = \{52\}$ and $D_1 = \{14.4\}$. Comparing these survival percentages with those given for the AJCC5 classification in Table 1, we obtain

$\Delta \bar{A} = 98 - 96.9 = 1.1$, $\Delta \bar{B} = 85 - 82.9 = 2.1$, $\Delta \bar{C} = 52 - 50.6 = 1.4$, $\Delta \bar{D} = 14.4 - 14.4 = 0.0$ and $\Delta \bar{P} = (98 + 85 + 52 + 14.4)/4 - (96.9 + 82.9 + 50.6 + 14.4)/4 = 62.35 - 61.2 = 1.15$. The positive figures here suggest that there has been an improvement in survival rate over time in Stages 1, 2 and 3, and over all stages. However, as this result takes no account of the change of classification system, we may be seeing an artefact of the Will Rogers phenomenon.

If we now apply our Will Rogers phenomenon correction method, we obtain $\widehat{\Delta A_0} = 98 - 96.9 = 1.1$, $\widehat{\Delta B_0} = 78.14 - 82.9 = -4.76$, $\widehat{\Delta C_0} = 50.58 - 50.6 = -0.02$ and $\widehat{\Delta D_0} = 16.51 - 14.4 = 2.11$.

The results $\Delta \bar{A} = 1.1$, $\Delta \bar{B} = 2.1$, $\Delta \bar{C} = 1.4$, $\Delta \bar{D} = 0$ erroneously suggest that the intervention has improved survival in Stages 1, 2 and 3, while having no effect on survival in Stage 4. However, when we correct for the Will Rogers phenomenon, we obtain a rather different result. The intervention had a positive effect on survival in Stages 1 and 4, a negative effect on survival in Stage 2, and no effect on survival in Stage 3. It is important to note here that the differences of value calculated are expressed in terms of the previous AJCC5 classification.

In the above calculations we can also take into account the uncertainty in the AJCC5 classification system percentage survival rate represented by the confidence intervals in Table 1. One general way of doing this is to simulate percentage survival rates from beta distributions with 0.025- and 0.975-quantiles given by these confidence intervals. We use $100,000$ simulations. We performed these simulations using R code (R Core Team, 2019) supplied by Blisle (2017) that we have verified. Here in brackets are associated 95% confidence intervals for the quantities that we compute above: $\Delta \bar{A} = 1.1 \ (-0.6, 3.3)$, $\Delta \bar{B} = 2.1 \ (0.1, 4.1)$, $\Delta \bar{C} = 1.4 \ (-2.9, 5.8)$, $\Delta \bar{D} = 0.0 \ (-3.8, 3.3)$ and $\widehat{\Delta A_0} = 1.1 \ (-0.6, 3.3)$, $\widehat{\Delta B_0} = -4.76 \ (-6.76, -2.76)$, $\widehat{\Delta C_0} = -0.02 \ (-4.32, 4.38)$ and $\widehat{\Delta D_0} = 2.11 \ (-1.69, 5.04)$. These intervals are consistent with the different conclusion obtained when we correct for the Will Rogers phenomenon that we discussed above, especially for Stage 2.

We have therefore described a method to correct for the artefactual changes that the Will Rogers phenomenon has introduced into our intervention study that has led to a less biased estimate of the effect of the intervention. There are, however, some important caveats associated with the analysis that we have presented above. Reducing the bias associated with the Will Rogers phenomenon is, for example, only one aspect of estimating cancer patient survival, meaning that the tool that we present may have associated with it a number of complex limitations. In general, in addition to statistical and mathematical techniques, a range of advanced clinical and biological insights are required to make use of cancer patient survival data; see Dickman and Adami (2006) and also Cox and Oakes (1984), Venables and Ripley (2002) Chapter 12, and Klein and Moeschberger (2003), the last of which discusses data from medicine, public health and epidemiology. In addition, it has been recognised for some time that a range of actions, including reducing delays in diagnosis and treatment through effective screening programmes, managing better older people with cancer, lifestyle interventions and collecting high quality data about treatment initiatives can lead to improvements in cancer survival rates; see Foot and Harrison (2011), for example.

## 4 Discussion

In this paper, we described the Will Rogers phenomenon and provided motivation for the need to correct for it from a range of epidemiological studies. We then outlined a method to correct for the Will Rogers phenomenon that depends on a specially constructed transition matrix. We provided worked biometrical examples of our method based on a dummy ecological study, migration economics and breast cancer epidemiology.

The Will Rogers phenomenon is recognised as one of the most important biases limiting the use of historical controls groups in experimental treatment trials (Miller et al., 2001; Sormani, 2009). Our method allows sensible inference about the effect of an intervention to be obtained in the presence of changes in diagnostic procedures. Avoiding the bias due to the Will Rogers phenomenon is also beneficial where historical control trials are preferred to placebo control trials, for example when non-treatment of a patient is unethical. Our method also allows for bias correction in long-term intervention studies and observational studies across changing diagnostic criteria. It may be especially appropriate in investigations in which there is a geographical or spatial component.

The calculations involved in our proposed method to correct for the Will Rogers phenomenon are only possible if all the information needed to classify new cases with reference to the previous classification systems is available. The cancer epidemiology example we discussed above provides a model for the sort of information that would be required. Without this information it may not be possible to find exactly the transition matrices that are fundamental to the calculations. Tan et al. (2015) is therefore an exemplar of the sort of study needed to provide the required information. In the absence of information about how classification definitions have changed, the transition matrix can be estimated from data (see Section 3.2 and also, for example, Bishop et al. (2007) and Stander et al. (1989) ) or defined using expert medical opinions, if they are available. Our methodology allows the user to specify a range of transition matrices and to assess their effect on the assessment of interventions.

In future work we plan to extend our methodology to correct for the Will Rogers phenomenon in the presence of covariates.

**Conflict of Interest**
*The authors have declared no conflict of interest.*

## References

AJCC Website. 2019. AJCC. [ONLINE] Available at: https://cancerstaging.org. [Accessed 27 June 2019].

Albertsen, P., *et al.* (2005) Prostate cancer and the Will Rogers phenomenon. *JNCI: Journal of the National Cancer Institute*, **97 (17)**, 1248–1253.

Bishop, Y. M. M., Fienberg, S. E., Holland, P. W. (2007). *Discrete multivariate analysis: Theory and practice*. New York: Springer

Blesch, K. S. (2005) Bias from stage migration in cancer survival. In *Encyclopedia of Biostatistics* (eds P. Armitage and T. Colton). Hoboken: Wiley.

Blisle, P. (2017) `beta.parms.from.quantiles`: an R function for computing beta distribution parameters. `http://www.medicine.mcgill.ca/epidemiology/joseph/pbelisle/BetaParmsFromQuantiles.html`. [Accessed 07 July 2019].

Borjas, G. J. (2017) The wage impact of the marielitos: A reappraisal. *ILR Review*, **70 (5)**, 1077–1110.

Clark, P. and Leaverton, P. (1994) Scientific and ethical issues in the use of placebo controls in clinical trials. *Annual Review of Public Health*, **15**, 19–38.

Cox, D. and Oakes, D. (1984) *Analysis of Survival Data*. Chapman and Hall, London.

Dickman, P. W. and Adami, H. (2006), Interpreting trends in cancer patient survival. *Journal of Internal Medicine*, 260: 103–117.

Docquier, F., Ozden, C. and Peri, G. (2011) The wage effects of immigration and emigration (English). Policy Research working paper: no. WPS 5556. *World Bank* `http://documents.worldbank.org/curated/en/675801468331816158/The-wage-effects-of-immigration-and-emigration`. [Accessed 07 July 2019].

Feinstein, A. R. (1979) Should placebo-controlled trials be abolished? *European Journal of Clinical Pharmacology*, **17(1)**, 1–4.

Feinstein, A. R., Sosin, D. M. and Wells, C. K. (1985) The Will Rogers phenomenon. Stage migration and new diagnostic techniques as a source of misleading statistics for survival in cancer. *The New England Journal of Medicine*, **312 (25)**, 1604–1608.

Felbermayr, G. J., Hiller, S. and Sala, D. (2010) Does immigration boost per capita income? *Economics Letters*, **107 (2)**, 177–179. `URL:http://opus.uni-hohenheim.de/volltexte/2009/331/pdf/300.pdf`. [Accessed 07 July 2019].

Fleming I.D., *et al.* (1997) *AJCC Cancer Staging Manual*. 5th ed. Philadelphia: Lippincott-Raven.

Fletcher, R. H. and Fletcher, S. W. (2005) *Clinical Epidemiology: The Essentials*. Baltimore: Lippincott Williams & Wilkins.

Foeglé, *et al.* (2005) Non-small-cell lung cancer in a French department (1982-1997): management and outcome. *British Journal of Cancer*, **92 (3)**, 459–466.

Foley, M. E., *et al.* (2013) Apparent improvement in survival for carcinoma of the cervix following the introduction of Chemoradiation – a Will Rogers phenomenon. *Irish Medical Journal*, **106 (3)**, 74–77.

Foot C. and Harrison T. (2011) *How to improve cancer survival: Explaining Englands relatively poor rates*. The Kings Fund: London

Gofrit, O. N., *et al.* (2008) The Will Rogers Phenomenon in urological oncology. *The Journal of Urology*, **179, 1**, 28–33.

Greene F.L., *et al.* (2002) *AJCC Cancer Staging Manual*. 6th ed. New York (NY): Springer-Verlag.

Grosclaude P. *et al.* (1995) Differences in treatment and survival rates of non-small-cell lung cancer in three regions of France. *British Journal of Cancer*, **72**, 1278–1282.

Klein, J. P. and Moeschberger, M. L. (2003) Survival Analysis: Techniques for Censored and Truncated Data. Second Edition. Springer.

Miller, K., Rahman, Z. and Sledge Jr., G. (2001) Selection bias in clinical trials. *Breast Disease*, **14, 1**, 31–40.

Ottaviano, G. I. and Peri, G. (2012) Rethinking the effect of immigration on wages. *Journal of the European Economic Association*, **10**, 152–197.

Qi, R. and Moul, J. (2017) African American men with low-risk prostate cancer *are* candidates for active surveillance: The Will-Rogers effect? *American Journal of Men's Health*, **11 (6)**, 1765–1771.

Quan, H., Zhang, B., Chuan-Stein, C., Jones, B., and on behalf of the EFSPI Integrated Data Analysis Efficacy Working Group (2017) Integrated data analysis for assessing treatment effect through combining information from all sources. *Statistics in Biopharmaceutical Research*, **9**, 52–64.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. `http://www.R-project.org/` [Accessed 07 July 2019].

Segen, J. C. (2012) Will Rogers Paradox. In *Segen's Medical Dictionary*. `https://medical-dictionary.thefreedictionary.com/Will+Rogers+paradox`. [Accessed 07 July 2019].

Sormani, M. P. (2009) The Will Rogers phenomenon: the effect of different diagnostic criteria. *Jounral of Neurological Science*, **287, Suppl 1: S46-9**.

Sormani M. P. and Bruzzi P. (2015) Can we measure long-term treatment effects in multiple sclerosis? *Nature Reviews Neurology*, **11**, 176–182.

Stander, J., Farrington D. P., Hill G., and Altham P. M. E. (1989) Markov chain analysis and specialization in criminal careers. *The British Journal of Criminology* 29(4): 317–35.

Tan, G.H., *et al.* (2015) The Will Rogers phenomenon in the staging of breast cancer — Does it matter? *Cancer Epidemiology*, **39 (1)**, 115–117.

Venables, W. N. and Ripley, B. D. (2002) *Modern Applied Statistics with S.* Fourth Edition. Springer, New York

Viele, K., *et al.* (2014) Use of historical control data for assessing treatment effects in clinical trials. *Pharmaceutical Statistics*, **13 (1)**, 41–54.