

Medium-term Mortality of Dutch Professional Soccer Players

Ruud H. Koning*
Remko Amelink*

Abstract

Professional athletes such as soccer players are expected to be in better health than the average member of their age cohort, if only because of the very demands imposed on them by their profession. Moreover, during their active career, their health is monitored on a regular basis. However, whenever a (former) well-known athlete dies, such an event is publicised widely. Perceived mortality among athletes may be distorted. In this article, we examine medium-term mortality of professional soccer players in The Netherlands. More specifically, we answer two questions. First, we examine if the mortality of soccer players is significantly different from mortality in the general population. Second, we assess whether mortality varies by team. We find that mortality among soccer players is significantly lower than in the general population, with no significant between teams variation. The methodology used in this article is easily extended to examine atypical mortality among other groups of athletes.

Keywords

Mortality; professional soccer; splines.

Introduction

Professional soccer players are expected to be in better health than the average member of their age cohort, because their profession demands a high level of physical fitness. Moreover, during their active career, their health is monitored on a regular basis by physicians (usually employed by the teams for whom they play). However, whenever a (former) well-known player dies, such an event is publicised widely. Three examples are the deaths of Daniel Jarque, Marc-Vivien Foe, and David di Tommaso. Daniel Jarque died on 8 August 2009. He was a player for Espanyol, a team in the Primera Division (the highest Spanish soccer league). Marc-Vivien Foe died from a fatal cardiac arrest in 2003 whilst playing a game for the Cameroon national team. He was playing for Manchester City at that time. In the Netherlands, FC Utrecht player David di Tommaso died unexpectedly of a cardiac arrest in 2005. Perceived mortality of soccer players may be distorted.

* University of Groningen, The Netherlands

In this article, we examine medium-term mortality of professional soccer players in The Netherlands. We consider a sample of all players active in the highest professional soccer league in The Netherlands (Eredivisie) in three seasons 1970/1, 1971/2, and 1972/3, and examine their status on 1 January 2009. More specifically, we address two main questions. First, are mortality rates of these soccer players significantly lower or higher than the ones in the Dutch population in general? Second, we examine whether there is any heterogeneity of mortality rates between teams. Even though it may be interesting to hypothesise on the causes of differences in mortality rates, if any, we are not able to attribute any deviations from expected mortality to causal factors. Causal explanation requires far more detailed data that are not presently available. We develop a methodology that can be generalised to other groups of athletes.

This study is related to and complements a number of other articles. Teramoto and Bungum (2010) survey fourteen articles on mortality and longevity of elite athletes. The common findings in the studies reviewed are that elite endurance athletes and mixed-sports athletes (soccer players belong to this category) survive longer than the general population. The likely primary cause of this effect is lower cardiovascular disease mortality. However, results on power athletes are mixed.

Sarna, Sahi, Koskenvuo, and Kaprio (1993) collected data on athletes representing Finland in elite international contests, and match athletes with healthy non-athletes of the same age and region using the Finnish Defence Forces conscription database. They find that Finnish team members have higher life expectancy that is mainly explained by decreased risk of cardiovascular mortality.

In the case of soccer, particular attention has been given to amyotrophic lateral sclerosis (ALS) as a cause of mortality (Belli and Vanacore 2005; Taioli 2007). Belli and Vanacore estimate standardised proportionate mortality ratios for a number of causes of death for 24,000 Italian soccer players active in the period 1960-1996. They find that the mortality ratios adhere substantially to expected mortality with the exception of mortality for diseases of the nervous system. In particular, ALS is more prevalent than expected, and this is possibly related to the use of dietary supplements and drugs.

Mortality among elite athletes as professional soccer players may differ from mortality in the general population for a number of reasons. First, athletes are healthy because they are a self-selected group of the population. Second, during their active career, their health status is monitored closely, and they have access to high quality medical care. Third, the soccer players we consider in this article may be relatively well-off after their career, enabling them to have better access to medical care after their career. Finally, we mention that being successful and having a high social status may also contribute to lower mortality outcomes (as has been shown for Academy Award winners by Redelmeier and Singh (2001)). On the other hand, mortality could be higher due to physical strain during the active career, possible use of performance enhancing drugs, or a celebrity life-style after (or possibly even during) the active career. Another reason for higher mortality could be depression or other mental health related disorders

as a result of poor adjustment to an alternative career or lifestyle. Given the growth in the size of the soccer industry, different mortality is potentially an important issue.

Due to the long time span between the active career and the moment of investigation, we need to allow for changes in mortality rates over time. Mortality rates have decreased for all ages, mainly due to improvements of medical knowledge and living conditions. In the analysis below, we take these changes into account.

Data and Methods

In this section, we present our research design in three steps. First, we discuss compilation of the dataset of professional soccer players. Using this dataset, we estimate expected mortality over time. Second, since these mortality rates change over time, we discuss how we allow for these rates to change over time. Third, we indicate how we compare expected mortality (based on the adjusted mortality rates) with observed mortality (as measured in the dataset with players).

We begin by describing how we collected the basic dataset of soccer players. We focus on all players who have played in the highest level of Dutch professional soccer ('Eredivisie') in one of the three seasons 1970–71, 1971–72, and 1972–73. For all players, we assess on 1 January 2009 whether he is living or dead (and the date of death). We choose players from these seasons for the following reasons. First, enough time should have elapsed between the active career and the moment of measurement of survival. If the moment of measurement is too soon after the active career, individual survival will be censored for too many players, making it hard if not impossible to distinguish between expected survival and observed survival. By 1 January 2009, 28 out of 371 players had died, survival in our sample is censored for 92.5 per cent of all observations. By considering three seasons instead of one season (for example, 1970–71), the number of observations is increased from 240 to 371. We use three consecutive seasons, so that all players have access to the same general state of medical knowledge, and have approximately similar stocks of health.

Lists of almost all players who were active in these three seasons were provided by the Koninklijke Nederlandse Voetbal Bond (the Dutch soccer association), and Infostrada (a private company specialising in collecting and publishing sports data). The list of Infostrada contained names of players who have played at least one match in one of the three seasons, and we use that list. Contract players with no match appearances are not representative of the population of top level professional soccer players and are not included. The list also contained the date of birth of each player, this information was checked against other sources, such as club websites and books. No discrepancies were found. To complete the data set, we collected information on mortality of each player using a variety of sources, websites of the teams being one of them. Finally, each team was sent a list with the names of the players, their dates of birth and dates of death if applicable, with a request to check these dates. We only use information that has been validated explicitly by the teams. Teams that did not respond, or were unable to check the dates of birth and death, are not included in the analysis.

Club	70-71		71-72		72-73		Overall	
	Total number of players	Number of deceased players	Total number of players	Number of deceased players	Total number of players	Number of deceased players	Total number of players	Number of deceased players
Ajax	15	3	18	0	19	0	23	3
AZ	23	1	0	0	22	2	32	2
FC Groningen	0	0	22	4	22	5	26	5
FC Twente	22	2	22	2	17	2	25	2
FC Utrecht	26	3	21	3	22	2	34	4
FC Volendam	24	0	26	0	0	0	31	0
Feyenoord	21	2	22	1	21	0	32	2
Haarlem	24	4	0	0	18	2	32	4
NAC Breda	21	0	19	0	20	1	30	1
NEC	22	0	24	0	22	0	38	0
PSV	17	0	19	0	18	0	26	0
Sparta	25	3	20	3	20	3	30	3
Vitesse	0	0	22	2	0	0	22	2
Total	240	18	235	16	221	17	371	28

Table 1: Number of players in portfolio. The total number of unique players of each club can be found in column 'Total'. Some players have played for multiple clubs in the dataset. Therefore, the total number in the dataset is smaller than the sum of the unique players of every club.

Table 1 presents the raw counts of deceased players by team. Thirteen out of 23 teams that played in the highest league in at least one of the three seasons responded. Some teams played only one or two seasons in the highest league, as for example Vitesse. The number of players in each pair of columns is not unique, so the last pair of columns is not the sum of the earlier columns. In total, 28 out of 371 players have died between their active career and 1 January 2009 (7.5 per cent).

In the second step, we need to estimate mortality in a group with this age composition. To do so, we need age and calendar time specific mortality rates. The Actuarial Association publishes mortality tables in The Netherlands for five year periods (for example, 2000–05), separately for men and women. Since we consider professional soccer players, we use the tables for men only. In the remainder, $p(x, t)$ denotes the probability that a male aged x survives one more year, in calendar year t . The corresponding mortality rate is $q(x, t) \equiv 1 - p(x, t)$. Survival probabilities have, in general, increased markedly over time (see for example figures A-1 and A-2 in the Appendix). The one year mortality rate of a 25-year old male has decreased from $q(25, 1970) = 0.0008582$ in 1970 to $q(25, 2010) = 0.0006914$ in 2010. Decreased mortality rates translate into increased residual expected lifetimes: for a 25-year old male residual lifetime has increased from 47.6 years in 1970 to 52.4 years in 2010. We smooth mortality rates over time using splines (Harrell 2001; Currie, Durban, and Eilers 2004) so as to avoid sudden discrete jumps in the one year mortality rates. If we were to base our analysis on constant mortality rates of 1971 (that is, we would not allow for improvements of medical science over time), we would underestimate survival and we would be biased towards concluding that mortality among soccer players is lower than mortality in the general population. In the last section, we discuss the sensitivity of our results to this approach.

Now that we have a portfolio of 371 soccer players of different ages, and one year mortality rates, we derive the probability distribution of the number of survivors at any moment in (calendar) time by simulation. We denote the number of players of age x at calendar time t by $N(x, t)$, so that the total number of players alive at calendar time t is $N(t) \equiv \sum_x N(x, t)$. In our model, we assume that all players are born on 1 January. Other than Dudink (1994), we do not reject the hypothesis that birthdates of players are uniformly distributed throughout the year. Let the probability that a player aged x alive at t survives for s years be denoted by $p(x, t, t + s)$ (so that one-year survival probability is $p(x, t) \equiv p(x, t, t + 1)$). This s -year survival probability is the product of one-year survival probabilities, taking aging and time dependence into account:

$$p(x, t, t + s) = p(x, t, t + 1) \times p(x + 1, t + 1, t + 2) \times \dots \times p(x + s - 1, t + s - 1, t + s) \tag{1}$$

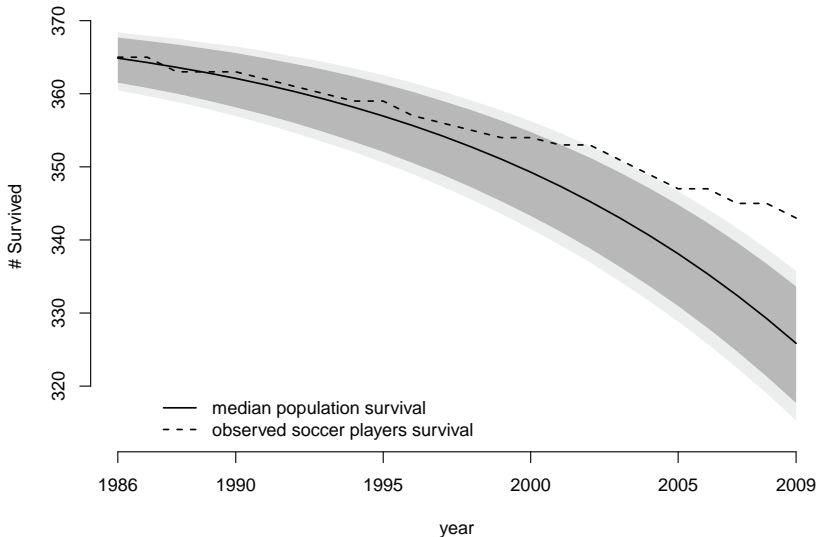
We fix t_0 , the starting point of the analysis to be 1 January 1971. The number of surviving players $N(x + s, t_0 + s)$ follows a binomial distribution with parameters $N(x, t_0)$ and $p(x, t_0, t_0 + s)$. The total number of players who survive at $t_0 + s$ is

$$N(t_0 + s) = \sum_{x+t_0+s} N(x + s, t_0 + s), \quad s = 1, 2, \dots, 38. \tag{2}$$

with the summation taking place over all relevant values of $x + s$. Since we consider three seasons, in our calculations we allow for an increase of our portfolio of players at times $t_0 + 1$ and $t_0 + 2$ (when the new players of 1971–72 and 1972–73 are added to the portfolio). The distribution of $N(t_0 + s)$ is not available in closed form, therefore we approximate it by simulating the individual terms in the sum in equation (2). This distribution gives survival over time if the pool of soccer players were to have similar mortality to the general population. We refer to this distribution as population survival.

As opposed to Belli and Vanacore (2005), Sarna et al. (1993), and Taioli (2007), we do not have access to administrative data. For this reason, we are unable to perform a case-control study, and we do not know the exact cause of death. While our dataset is less rich in this respect, our approach is flexible and can easily be extended to other settings.

Figure 1: Survival curve portfolio of all players, dark grey area: 80% confidence interval, light grey area: 90% confidence interval



Results

Mortality per team is given in Table 1. We have been able to obtain complete information on 371 unique players, 28 of whom died before 1 January 2009. The first question is how observed mortality compares to expected mortality, using the approach discussed in the previous section. First, in Figure 1 we graph median survival over time and its associated 80 per cent and 90 per cent confidence bands. The confidence bands are obtained by linear interpolation. The solid decreasing line in the center of the bands shows the expected development of the number of players in the portfolio, if the general population mortality rates were to apply to them. As discussed in the previous section, this curve does

Year	Whole portfolio					p-value
	$n(t_0)$	$n(t)$	$\Pr(N(t) < n(t))$	$\Pr(N(t) = n(t))$	$\Pr(N(t) > n(t))$	
2000	371	354	0.805	0.059	0.136	0.195
2001	371	353	0.849	0.048	0.103	0.151
2002	371	353	0.927	0.027	0.046	0.073
2003	371	351	0.925	0.026	0.048	0.075
2004	371	349	0.928	0.024	0.047	0.072
2005	371	347	0.936	0.022	0.043	0.064
2006	371	347	0.977	0.009	0.014	0.023
2007	371	345	0.982	0.007	0.011	0.018
2008	371	345	0.995	0.002	0.003	0.005
2009	371	343	0.997	0.001	0.002	0.003

$\alpha = 0.05$
 $\alpha = 0.10$
 $\alpha > 0.10$

Table 2: Development of complete portfolio over time, with p-values by year.

reflect decreasing mortality rates over time. The dashed line indicates observed survival, and after 2001 observed survival is higher than the 80 per cent upper confidence limit. By 1 January 2009, observed survival is noticeably higher than the 90 per cent upper confidence limit. We have calculated the exact p -value in 2009 to be 0.003. In Table 2 we give observed values for observed survival $n(t)$ for the period 2000–09. Also, we show how population survival is distributed around this point. The column p -value equals $\Pr(N(t) \geq n(t))$, the lowest level of significance that would give a rejection of the null hypothesis in year t . Note that this is a one-sided p -value, the hypothesis being tested being that observed survival exceeds population survival. As of 2006 the p -value is smaller than the level of significance $\alpha = 0.05$. Observed survival is significantly higher than survival in the general population, in other words, mortality among soccer players is lower than the mortality in the general population. In the Appendix, we give similar tables for survival of players from each of the three seasons, and these tables lead to the same conclusion.

The second question is whether mortality varies by team. The players represented thirteen different teams. In total, we have information on 35 team-years out of 54 possible teams-years (eighteen teams times three seasons). The aggregate mortality rate is 7.5 per cent over 37 years (so the average mortality rate is 0.27 per cent annually). The number of deaths per team seems to vary, but the number of players at risk differs by team as well. To assess whether or not this variation by team is systematic, we calculate Pearson's χ^2 -statistic, which is 16.2 in this case. Under the null hypothesis of equal mortality between teams, we calculate the p -value as 0.12, by permutating all deaths and survivals over the teams. We prefer a permutation test to an asymptotic test because of the limited number of observations and small mortality rate. The hypothesis of equal mortality between teams cannot be rejected.

Discussion

In this article, we have shown that the medium-term mortality in a sample of professional soccer players active in the highest league of Dutch soccer (1970–71 until 1972–73) is significantly lower than mortality in the general population ($p = 0.003$). We do take into account that mortality rates decrease over time. The difference between expected survival and survival in the portfolio of soccer players has been increasing since 2006. Also, we have shown that mortality does not vary significantly by team.

In our analysis, we smoothed mortality rates over time. We performed the same test as presented in the last column of Table 1 under two alternative models. First, we did not smooth mortality rates, but we used the five-year mortality rates as published by the Actuarial Association. That is, we allowed for mortality rates that change over time, but in a discrete, discontinuous fashion. The p -value in 2009 is in that case 0.003 as well. The second alternative scenario is to ignore changes in mortality and assume that the mortality rates of 1971 apply to the whole 1970–2009 period. In that case, the p -value in 2009 is 0.000, smaller than 0.003. The difference between this approach and our approach is perhaps best illustrated by looking at survival by 1 January 2000. Allowing for non-constant

mortality rates, the p -value is 0.195, under the incorrect assumption of constant mortality rates, it is $p = 0.004$. In other words, expected survival is underestimated by assuming mortality rates that are constant over time.

As a final remark, we want to point out that the methodology developed in this article can be easily applied to other sports to assess development of mortality patterns over time. Data requirements of the approach are modest. In particular, this approach can be used to shed more light on the inconsistent results found for athletes participating in anaerobic (power) sports (see for example, in Teramoto and Bungum 2010).

Acknowledgements

We thank Infostrada, KNVB, the Actuarial Institute, and Henk Grim for helping compiling the datasets. We also thank the football teams that were able to check our data. This research is part of the Research program Passion, Practice and Profit. The program ran from 2007 to 2010 and was financed by the Dutch Ministry of Health, Welfare and Sport (VWS). It was coordinated by the W. J. H. Mulier Institute for Social Science Sportsresearch. In the program researchers of the institute joined forces with researchers from four associated universities (Tilburg University, University of Amsterdam, Utrecht University and University of Groningen).

References

- Belli, S. and Vanacore, N. (2005) 'Proportionate mortality of Italian soccer players: Is amyotrophic lateral sclerosis an occupational disease?', *European Journal of Epidemiology*, 20(3), pp. 237–242.
- Currie, I., Durban, M. and Eilers, P. (2004) 'Smoothing and forecasting mortality rates', *Statistical Modelling*, 4(4), pp. 279–298.
- Dudink, A. (1994) 'Birth date and sporting success', *Nature*, 368, p. 592.
- Harrell, Jr, F. E. (2001) *Regression Modelling Strategies*, Springer, New York.
- Redelmeier, D. A. and Singh, S. M. (2001) 'Survival in Academy Award-winning actors and actresses', *Annals of Internal Medicine*, 134(10), pp. 955–962.
- Sarna, S., Sahi, T., Koskenvuo, M. and Kaprio, J. (1993) 'Increased life expectancy of world class male athletes', *Medicine and Science in Sports and Exercise*, 25(2), pp. 237–244.
- Taioli, E. (2007) 'All causes mortality in male professional soccer players', *European Journal of Public Health*, 17(6), pp. 600–604.
- Teramoto, M. and Bungum, T. J. (2010) 'Mortality and longevity of elite athletes', *Journal of Science and Medicine in Sport*, 13, pp. 410–416.

About the Authors

- » **Professor Ruud H. Koning** is professor of Sports Economics at the University of Groningen, The Netherlands. His research fields include economics and econometrics of sports, applied microeconometrics and applied actuarial science. He is head of the Department of Economics, Econometrics, and Fi-

nance of the Faculty of Economics and Business. Besides, he is member of the supervisory board of Algemeen Belang (funeral insurer), and avid soccer fan. He can be reached at r.h.koning@rug.nl.

- » **Remko Amelink** a senior actuarial analyst at ABN AMRO Insurance. Currently, he is following the Executive Master of Actuarial Science program at TiasNimbas (Tilburg, The Netherlands) to become a certified actuary. His email address is amelinkr@hotmail.com.

Appendix

In figures A-1 and A-2, we give two examples how one-year survival probabilities change over time, for two different ages, $x = 24$ and $x = 65$. The dots indicate the probabilities as published by the Actuarial Association. As discussed in the text, the Actuarial Association publishes tables in five year intervals. The smooth curve is the spline we used to smooth the survival rates. Smoothing was obtained by estimating cubic B -splines with two knots, one at 1981 and one knot at 1995. These spline functions have a continuous first and second derivative. By using such smooth functions, we implicitly assume that survival (or its counterpart, mortality) changes gradually over time.

Figure A-1: Smoothed one-year survival probabilities over time, age $x = 24$

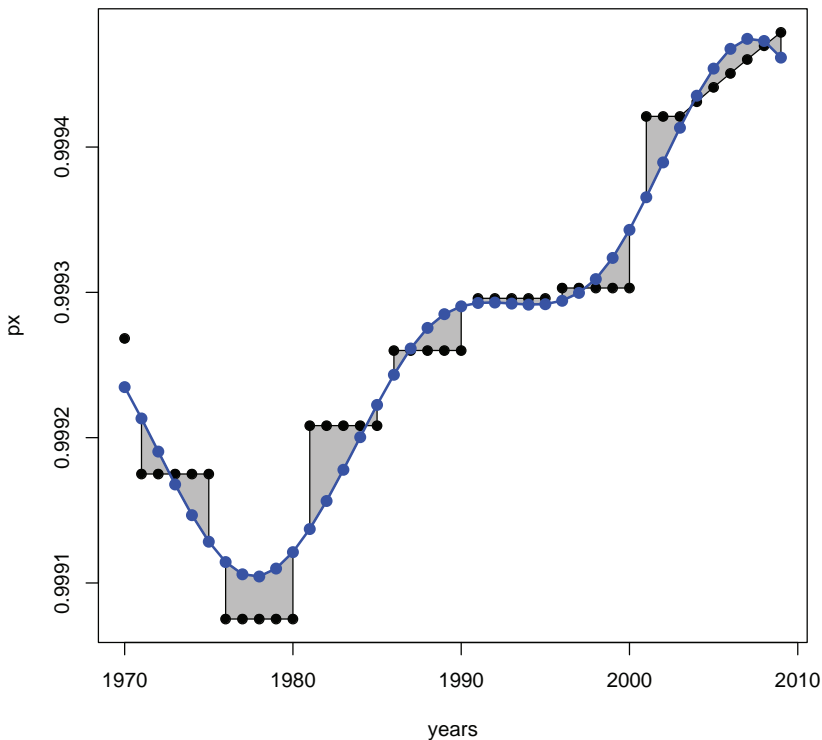


Figure A-2: Smoothed one-year survival probabilities over time, age $x = 65$

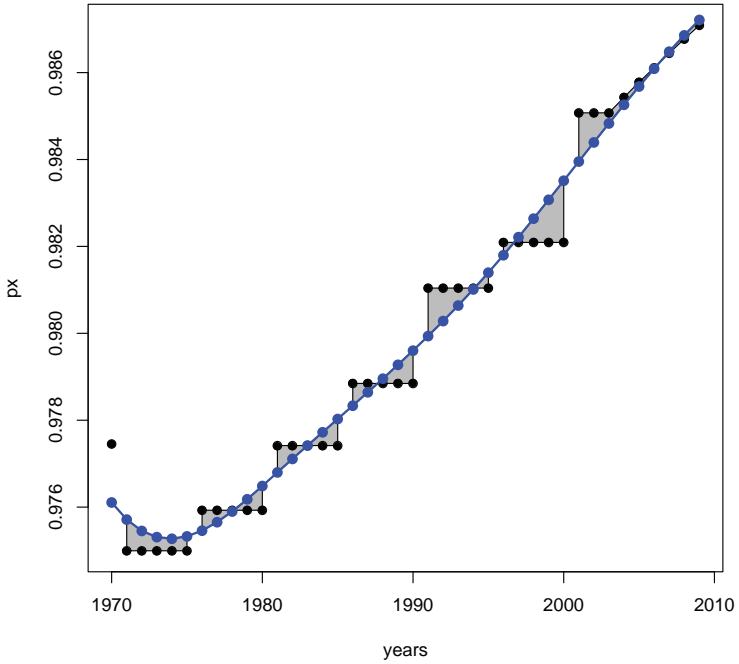
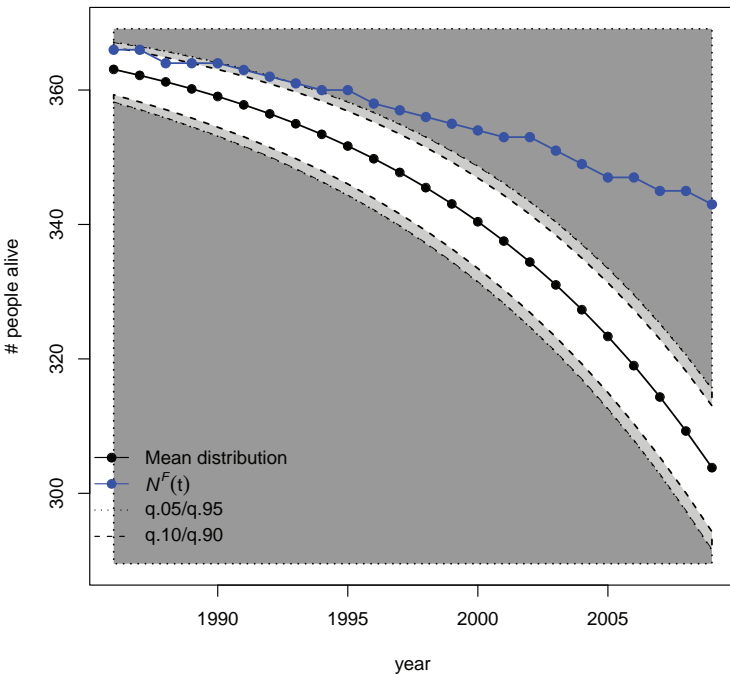


Figure A-3: Survival curve of portfolio of all players when mortality is assumed to be constant at rates of 1971



Year	Portfolio season 1970/71					
	$n(t_0)$	$n(t)$	$\Pr(N(t) < n(t))$	$\Pr(N(t) = n(t))$	$\Pr(N(t) > n(t))$	p -value
2000	240	228	0.709	0.090	0.201	0.291
2001	240	228	0.814	0.066	0.120	0.186
2002	240	228	0.893	0.043	0.064	0.107
2003	240	227	0.909	0.037	0.055	0.091
2004	240	227	0.956	0.020	0.025	0.044
2005	240	225	0.942	0.024	0.034	0.058
2006	240	225	0.976	0.011	0.013	0.024
2007	240	223	0.971	0.012	0.016	0.029
2008	240	223	0.990	0.005	0.005	0.010
2009	240	222	0.994	0.003	0.003	0.006

Table A-1: p -values of null-hypothesis, $t = 2009$, players 1970/71.

Year	Portfolio season 1971/72					
	$n(t_0)$	$n(t)$	$\Pr(N(t) < n(t))$	$\Pr(N(t) = n(t))$	$\Pr(N(t) > n(t))$	p -value
2000	235	225	0.812	0.072	0.116	0.188
2001	235	224	0.820	0.067	0.113	0.180
2002	235	224	0.894	0.044	0.062	0.106
2003	235	223	0.906	0.039	0.055	0.094
2004	235	221	0.873	0.047	0.081	0.127
2005	235	220	0.895	0.039	0.066	0.105
2006	235	220	0.950	0.020	0.030	0.050
2007	235	219	0.963	0.016	0.021	0.037
2008	235	219	0.986	0.007	0.007	0.014
2009	235	219	0.995	0.002	0.002	0.005

Table A-2: p -values of null-hypothesis, $t = 2009$, players 1971/72.

Year	Portfolio season 1972/73					<i>p</i> -value
	$n(t_0)$	$n(t)$	$\Pr(N(t) < n(t))$	$\Pr(N(t) = n(t))$	$\Pr(N(t) > n(t))$	
2000	221	211	0.663	0.106	0.231	0.337
2001	221	210	0.665	0.102	0.234	0.335
2002	221	210	0.771	0.080	0.149	0.229
2003	221	209	0.783	0.074	0.143	0.217
2004	221	208	0.800	0.068	0.132	0.200
2005	221	206	0.747	0.076	0.178	0.253
2006	221	206	0.848	0.053	0.100	0.152
2007	221	205	0.875	0.044	0.081	0.125
2008	221	205	0.937	0.025	0.038	0.063
2009	221	204	0.954	0.019	0.027	0.046

$\alpha = 0.05$  $\alpha = 0.10$  $\alpha > 0.10$ 

Table A-3: *p*-values of null-hypothesis, $t = 2009$, players 1972/73.