Explaining Patterns of Age-Specific Performance

Maren Rebke

PhD Thesis
2011

Department of Biological Sciences
Faculty of Life-Sciences
Imperial College London

Produced in collaboration with the Max Planck Institute for Demographic Research and the Institute of Avian Research ‘Vogelwarte Helgoland’.
The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without the prior written consent of the author.
Declaration

This thesis is my own work.

Co-authors, where appropriate, are listed on chapters that are published, submitted or in preparation for publication. If a co-author is listed first, I am a contributing author and I did not lead the research. People who provided less formal advice are named in the acknowledgments.
Individual life histories are frequently studied to gain insight into the mechanisms of ageing. However, various challenges complicate the accurate quantification of age-specific variation in fitness. In this thesis I develop and apply methods to accurately characterise patterns of ageing, and to explain why such patterns arise.

All mammals and birds have an upper bound on litter size, and for many species this limit is quite low. In addition, in many species, not all individuals breed at every possible opportunity. Reproduction should consequently be considered as two processes: whether an individual breeds or not and the number of offspring produced. These processes mean that reproduction in many species does not follow a Poisson process as is often assumed in analyses of breeding performance. A more appropriate model for a repeated ordinal response like annual reproductive success is a proportional odds model with a random intercept for individuals. Such a model has not previously been used in ecology or evolutionary biology. I apply this model to analyse age and temporal variation in the number of fledglings produced annually by male and female common terns (*Sterna hirundo*). I use data collected from this intensively studied, long-lived species, repeatedly throughout the thesis.

The proportional odds analysis reveal that reproductive performance in females initially increased with age, before declining as individuals began to senesce. But why does this pattern arise? Is it purely an effect of getting physiologically older or are other processes involved? I estimate the effect of the length of time spent with the current partner using the common tern data. Despite the quality of the data, it is not always obvious if unmarked partners are new or not. I use a hierarchical Bayesian model of the steps that lead to the number of fledglings. Modelling this complicated process requires a complex model, but results show that no substantial amount of observed age-related patterns in reproductive performance can be attributed to length of pair bond.

While the proportional odds and Bayesian analyses account for repeated measures on individuals they do not account for compositional change. Such a change in
the composition of the population caused by heterogeneity between individuals can mask true rates of individual change. I develop a novel retrospective decomposition method related to the Price equation to address this issue. The equation gives the exact contributions of selective disappearance and average change in individual performance among survivors to the aggregate change at the level of the population. This equation can be extended by including a term for the compositional change due to selective appearance of individuals in the study population.

I apply this decomposition to the common tern dataset to disentangle whether apparent increases and decreases in reproductive performance with age reflect genuine changes within individuals or are an artefact of compositional change in a heterogeneous population. I show an improvement in average reproductive performance of individuals over most of adult life and give support for reproductive senescence at old ages. I show that the contribution of compositional change is of minor importance, suggesting that population-level averages accurately capture the individual-level ageing process well.

Can the decomposition method I develop be applied to other systems? Does it lead to similar conclusions? I apply it to two different datasets dealing with functioning at old age in humans: the ability to live independently in the Danish 1905-cohort, and cognitive functioning for people aged 80 and older participating in the Chinese Longitudinal Health and Longevity Survey. In both studies I reveal that average individual functioning declines at old ages. Although the decline is also apparent at the population level it is less strong due to the tendency of individuals with lower functioning to drop out earlier. Finally, I illustrate the general use of the decomposition by applying it to epidemiological and economic studies in the appendix.

Overall, I find that reproductive performance improves over many age classes before senescence begins. Numerous processes can influence rates of age-related change, with results apparently specific to the trait and population under study.
First and foremost, I would like to thank my supervisors Tim Coulson and Jim Vaupel for their advice and support. My thesis benefitted greatly from getting the perspectives of two researchers from different fields.

I thank Tim especially for coming up with many good research ideas, for always giving very valuable comments and always answering all my questions instantly – I do not know anybody who replies to emails as quickly as Tim does. Further, I thank him and the entire ‘SLAPPED’ group as well as the Silwood bar crowd for always making me feel welcome when I stayed in Silwood Park. In particular, I would like to thank Kevin and also Audrey Wallace for hosting me even in Zambia and making the visit there such an unforgettable experience.

I am grateful to Jim especially for bringing me into demography and thank him for very valuable discussions on different aspects of my research as well as always being greatly supportive and enthusiastic. I am grateful for the financial support provided by the Max Planck Institute for Demographic Research (MPIDR). The inspiring atmosphere and excellent working conditions at the MPIDR made working there a great experience. I would like to thank Gunde Paetrow and Ulrike Liebich for their big contribution to creating these excellent working conditions, the computer support team for always fixing my computer or software problems and Rainer Heuer for solving organisational issues related to my funding. I am grateful to the ‘Evolutionary Biodemography’ and the ‘Modelling the Evolution of Ageing’ groups for many discussions and inspiring meetings. I thank David Thomson for his input into the development of the overall research topic and his advice at the beginning of my PhD-time.

Special thanks also to my additional advisor Peter Becker. He provided the fantastic common tern dataset and invaluable insights about this species as well as other helpful comments. Nobody knows the common tern better than he does. I am especially grateful to Peter Becker and all the fieldworkers and technical assistants who contributed to the large dataset. Additionally, I would like to thank Peter
Acknowledgements

Becker for giving me the chance to see how the data is collected. He and his group generously made my visits to the Institute of Avian Research very enjoyable.

This thesis would not have been possible without the invaluable input of several co-authors to whom I am deeply thankful. I wish to thank Fernando Colchero for bringing me into Bayesian statistics, for investing so much time in programming our model together, for being always super supportive and always available when I needed his help. I would like to thank Jutta Gampe for her invaluable statistical advice and for always making some time for me in her busy schedule. Even when I asked at the very last minute before a deadline she never refused to help me. I am grateful to Zhen Zhang for having the idea and giving me the opportunity to work on the Chinese Longitudinal Health and Longevity Survey (CLHLS) and for multiple thoughtful discussions. Further, I wish to thank Jim Oeppen who always provided great discussion, never hesitated to help me if I had any questions or concerns, gave me lots of advice and came up with many good ideas for applications of the decomposition method.

My work has greatly benefitted by comments, suggestions and discussions with Annette Baudisch, Emmanuelle Cam, Vladimir Camudas-Romo, Hal Caswell, Dan Nussey, Anna Oksuzyan, Jay Rotella, Alexander Scheuerlein, Uli Steiner, Shripad Tuljapurkar, Kenneth Wachter and numerous guests at the MPIDR as well as participants of the various conferences I have attended. I would like to thank Shripad Tuljapurkar and his group for being such nice hosts during my visit in Stanford.

I thank Rune Haubo Christensen for patiently answering all my questions concerning his R-package ordinal. I am indebted to Giancarlo Camarda for being my statistical support team whenever I needed help. I would also like to thank Iris Levi-tis for copy editing my Germanised English and working through the entire thesis in a very short amount of time.

I am grateful to Kaare Christensen for making the summary data of the Danish 1905-cohort available in his publication, to Yi Zeng and his colleagues for conducting the CLHLS and providing part of the dataset to us and to the authors of the datasets used in the appendix for making them publicly available.

Further, I would like to thank all my friends from inside and outside the institute, who always reassured me and believed that I could do it, for the distructions. I thank Alyson van Raalte in particular for bearing to stay in an office with me for such a long time, correcting various English texts I have written, establishing the bread-club and always bringing me the most useful small presents from wherever she goes. Further, I thank all the current and previous members of the one o’clock lunch group for always making the lunch break so enjoyable and Adam Lenart for
small tea breaks that helped me to survive the hard working days.

I am extremely grateful to my parents for their support throughout my life. I would also like to thank my aunt and uncle who always believed in me – their survival kit helped me over the last period of my PhD.

Last, but not least, my special thanks to Christian Drücker who is much nicer to me than I deserve. I thank him for always staying by me and supporting me wherever he can without ever complaining. Without him the formatting of my thesis would have driven me to despair and I would not have survived the stressful ending phase of my PhD.
Contents

Declaration 3
Abstract 5
Acknowledgements 7
Contents 11
List of Figures 15
List of Tables 23

1. Introduction 25
1.1. Theoretical Background ............................................. 26
  1.1.1. Classic Evolutionary Theory of Ageing .......................... 26
  1.1.2. Recent Models of the Evolutionary Theory of Ageing and Life
         History Theory Models ........................................... 28
1.2. Evidence From the Wild ........................................... 29
1.3. Challenges Addressed in this Thesis .............................. 30
  1.3.1. Appropriate Model for Reproductive Data of Long-Lived Ver-
         tebrates .......................................................... 30
  1.3.2. Pair Bond Length ............................................. 31
  1.3.3. Heterogeneity ................................................ 31
1.4. Datasets ................................................................ 33
  1.4.1. Common Tern .................................................. 33
  1.4.2. Danish 1905-Cohort and the Chinese Longitudinal Health and
         Longevity Survey .................................................. 34
  1.4.3. Epidemiological and Economic Datasets ......................... 35
1.5. This Thesis .......................................................... 35
2. Interactive Effects of Age and Environmental Conditions on Reproductive Performance in a Long-Lived Bird 39

2.1. Summary ................................................................. 40

2.2. Introduction ............................................................. 40
   2.2.1. Models to Study Age-Specific Reproductive Performance ... 41
   2.2.2. Appropriate Model ............................................. 42
   2.2.3. Outline ......................................................... 44

2.3. Methods ................................................................. 45
   2.3.1. Study Details .................................................. 45
   2.3.2. Model Details ................................................ 46

2.4. Results ................................................................. 49
   2.4.1. Proportional Odds Model with Random Effects .............. 49
   2.4.2. Comparison with a Poisson Model ............................ 52

2.5. Discussion ............................................................... 53
   2.5.1. Our Results in Comparison to Previous Findings .......... 53
   2.5.2. Comparison With a Poisson Model .......................... 61
   2.5.3. Potential Future Extensions of the Model .................. 61
   2.5.4. Conclusion .................................................... 61

3. Better the Devil You Know:
   Common Terns Stay With a Familiar Partner Although Pair Duration
   Does Not Affect Breeding Output 63

3.1. Summary ................................................................. 64

3.2. Introduction ............................................................. 64

3.3. Methods ................................................................. 69
   3.3.1. Study Details .................................................. 69
   3.3.2. Assignment of the Partner Status ............................ 72
   3.3.3. Model Details ................................................ 72
   3.3.4. Bayesian Modelling ......................................... 81

3.4. Results ................................................................. 82

3.5. Discussion ............................................................... 86
   3.5.1. The Model ..................................................... 86
   3.5.2. Our Results in Comparison to Previous Findings .......... 88
   3.5.3. Discussion of Model Details ................................. 95
   3.5.4. Implications of the Main Findings .......................... 96

3.6. Supporting Information ............................................... 98
7.2.2. Comparative Study ........................................... 158
7.2.3. Long-Term Effects of Partner Change ................. 158
7.3. Conclusion .................................................... 159

A. Exploring Longitudinal Data Subject to Non-Monotone Missingness:
   a Simple Decomposition of Changes .......................... 161
   A.1. Summary .................................................... 162
   A.2. Introduction ............................................... 162
   A.3. Methods .................................................... 164
       A.3.1. Data Sources ......................................... 165
   A.4. Examples ................................................... 166
       A.4.1. Example 1: The Impact of German Medical Funding Reform
               on Doctor Visits – a Two Panel Study .................. 166
       A.4.2. Example 2: Schizophrenia and Informative Censoring ........ 168
       A.4.3. Example 3: The Muscatine Childhood Obesity Study ........... 171
       A.4.4. Example 4: The Relationship Between Unemployment and
                Absenteeism in a German Company .................... 173
   A.5. Discussion .................................................. 173

Bibliography ....................................................... 177
List of Figures

1.1. An adult common tern with chick. Photo credit: Maren Rebke. . . . 34

2.1. The linear predictor for good (blue dotted line), medium (black line) and bad years (red dashed line) for females (left) and males (right). . 49

2.2. Probability of being in one of the following breeding categories over age: not breeding (dark blue), breeding and 0 fledglings (light blue), breeding and 1 fledgling (red), breeding and 2 fledglings (orange), breeding and 3 fledglings (yellow). The results are given for females and males separately in bad, medium and good years. . . . . . . . . . . . . . 51

2.3. The average number of fledglings for good (blue dotted line), medium (black line) and bad years (red dashed line) for females (top) and males (bottom) predicted by the proportional odds (left) and the Poisson model (right) with random effects. . . . . . . . . . . . . . . . . . . . . . . . 53

3.1. The structure of the model. It shows the hierarchical steps that lead to the number of fledglings, where the letters label the steps referred to in the text (a – survival analysis and presence in the colony, b – partner choice, c – number of fledglings per breeding attempt, d – additional breeding attempts). The parts of the model without a coloured background are processes at the single individual level. The yellow background highlights the part of the model that depends on the decision of both partners. The blue background represents parts that are modelled at the pair level. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 70
3.2. Flowchart of the assignment of partner status: 0 – new partner, 1 – previous partner, 2 – unknown if new or previous partner. In order to assign the partner status the following questions are asked: Is the 1st breeding time and thus the entire breeding history known? Is the current partner marked or unmarked? Is the current breeding event the bird’s first time breeding? Then the current partner ID is compared to the list of all previous partners. At the end certain assumptions are made to simplify the model (the parts in red).

3.3. The survival (left) and mortality function (right). The curves start at age two, marked by the light grey line, as described in section 3.3.3. The grey area is the 95% predictive interval, the white line in the right hand panel is the predicted curve using the average estimated parameters.

3.4. Function of the partner choice. In this example the bird has a choice between two partners, partner A and partner B. The graphs show the probability of breeding with partner B dependent on the number of times the bird has bred with that partner in the past, given that it has never bred with partner A (left) or has bred twice with partner A in the past (right). The grey area is the 95% predictive interval and the white line the predicted curve using the average estimated parameters.

3.5. A contour plot of the number of fledglings per pair and breeding attempt dependent on the age of the male and the female and their quality for the model with individual effects (left) and without (right). The quality of the members of a pair is here defined in terms of breeding performance, where good is an individual effect of 0, medium of -0.5 and bad of -1.

3.6. The shape of the link function (left) for the number of fledglings over age for males and the corresponding functions of the actual number of fledglings (right) depending on the quality of the individuals. Quality is here defined in terms of reproductive output. It enters the equation as a shift of the intercept in the link function, where the best quality corresponds to an individual effect of \( u = 0 \) and the worst quality to \( u = -1 \).

3.7. The probability of breeding a second (left) and a third time (right) in one breeding season dependent on the breeding success on the previous occasion. The error bars represent 95% predictive intervals.
3.8. Correlation of ages (left) and individual effects (right) of both partners within a pair when both partners are marked. When the same pair of partners appears more than once only the first time it appears is used for the analysis. The grey lines were estimated by linear regression. The coefficients of determination are $R^2 = 0.2704$ for the correlation of ages and $R^2 = 0.297$ for the correlation of qualities. The curves are estimated with an intercept of 3.0021 and slope of 0.4615 for the former and an intercept of 0.2857 and slope of 0.5297 for the latter. ................................................................. 99

3.9. Correlation of age at death and the individual effect for birds with transponder. The grey line was estimated by linear regression. The coefficient of determination is $R^2 = 0.0264$, the intercept was estimated with 2.2472 and the slope with 8.3902. .................................................. 99

5.1. Graphical representation of decomposition of the aggregate change at the population level into average ontogenetic development and change due to selective disappearance. $P$ is the aggregate change in the mean number of fledglings produced from one age to the next – in this specific case we use data between age 5 and age 6, $s$ is the average ontogenetic development and $d$ denotes the change due to selective disappearance. $V_5$ and $V_6$ are the averages of the number of fledglings at age 5 and 6 respectively and $v_5$ and $v_6$ are the corresponding averages for the survivors. ................................................................. 114

5.2. Reproductive patterns over age (959 individuals with 4307 observations). Reproduction is measured by mean relative reproduction. The dark blue dots depict the observed pattern for the population. The light blue triangles show the pattern that would result from average ontogenetic development; the value at any age is the sum of average changes among survivors up to that age. Similarly, the orange squares illustrate the pattern that would result if there was only change due to selective disappearance and the yellow diamonds represent the pattern if the only change in relative reproduction was caused by selective appearance. ................................................................. 118
5.3. Change in relative reproduction from age $x$ to $x + 1$ averaged over all individuals and all ages. (A) Decomposition of population change into average ontogenetic development, selective disappearance and selective appearance in percent (959 individuals with 4307 observations). (B) Change in relative reproduction for the age groups 3-13 and 14-20, respectively, at the population level (dark blue bars), and the contributions to this change due to average ontogenetic development (light blue bars) and selective disappearance (orange bars; 955 individuals with 4303 observations). In the old age group the average individual decrease in reproductive performance is also apparent at the population level but to a smaller extent because of the slight tendency of less reproductive individuals to suffer higher mortality. The error bars represent 95% confidence intervals.

5.4. Graphical representation of Equation 5.16 to decompose the aggregate change at the population level into average ontogenetic development, change due to selective disappearance, and change due to selective appearance when newly entering individuals have a higher mean than the surviving individuals already in the study (A) and newly entering individuals have a lower mean than the surviving individuals (B). $P$ is the aggregate change in the considered characteristic from age $x$ to the age $x + k$, $s$ is the average ontogenetic development, $d$ denotes the change due to selective disappearance, and $a$ the change due to selective appearance. $V_x$ and $V_{x+k}$ are the averages of the characteristic on the observed population level at age $x$ and $x + k$, respectively, and $v_x$ and $v_{x+k}$ are the corresponding averages for the survivors.
5.5. Change in number of fledglings from age $x$ to $x + 1$ averaged over all individuals and all ages. (A) Decomposition of population change into average ontogenetic development, selective disappearance and selective appearance in percent (959 individuals with 4307 observations). The proportional change is equivalent to an absolute mean change from one age to the next of 0.091 fledglings at the population level, a positive average ontogenetic development of 0.065; a change of 0.015 is due to selective disappearance and 0.01 to selective appearance. (B) Change in number of fledglings for the age groups 3-13 and 14-20, respectively, at the population level (dark blue bars), and the contributions to this change due to average ontogenetic development (light blue bars) and selective disappearance (orange bars; 955 individuals with 4303 observations). The error bars represent 95% confidence intervals.

5.6. Change in relative reproduction from age $x$ to $x + 1$ at the population level (dark blue bars), and the decomposition of this change into average ontogenetic development (light blue bars) and selective disappearance (orange bars; 955 individuals with 4303 observations). The error bars represent 95% confidence intervals.

5.7. Change in relative reproduction from age $x$ to $x + 1$ averaged over all females and all ages. (A) Decomposition of population change into average ontogenetic development, selective disappearance, and selective appearance in percent (411 individuals with 1893 observations). (B) Change in relative reproduction for the age groups 3-13 and 14-20, respectively, at the population level (dark blue bars) and the contributions to this change due to average ontogenetic development (light blue bars) and selective disappearance (orange bars; 408 individuals with 1890 observations). The error bars represent 95% confidence intervals.
5.8. Change in relative reproduction from age $x$ to $x+1$ averaged over all males and all ages. (A) Decomposition of population change into average ontogenetic development, selective disappearance and selective appearance in percent (468 individuals with 2123 observations). The average change among survivors contribution can be higher than 100% if the changes due to selective disappearance and/or selective appearance go in the opposite direction. (B) Change in relative reproduction for the age groups 3-13 and 14-20, respectively, at the population level (dark blue bars) and the contributions to this change due to average ontogenetic development (light blue bars) and selective disappearance (orange bars; 467 individuals with 2122 observations). The error bars represent 95% confidence intervals. ........................................ 132

6.1. Change in percentages of the Danish 1905-cohort classified as independent between four assessments from 1998 to 2005 which correspond to the ages 92/93 and 99/100 for the considered cohort. Males and females are combined. The dark blue bars show the change at the aggregate population level, the light blue bars the change in independence among the surviving individuals, and the orange bars represent the change that is explained by selective disappearance. ................ 141

6.2. Change in MMSE score for the age group 80-89 at the population level (dark blue bars) and the contributions to this change due to average individual change (light blue bars) and compositional change due to selective disappearance (orange bars) and new appearance of individuals in the study (yellow bars). The number of individuals considered in the three change intervals is 6026, 6045 and 4198 respectively. The error bars represent 95% confidence intervals. ......................... 143

6.3. Change in MMSE score for the age group 90-99 at the population level (dark blue bars) and the contributions to this change due to average individual change (light blue bars) and compositional change due to selective disappearance (orange bars) and new appearance of individuals in the study (yellow bars). The number of individuals considered in the three change intervals is 4888, 5501 and 3582 respectively. The error bars represent 95% confidence intervals. ......................... 144
6.4. Change in MMSE score for the age group +100 at the population level (dark blue bars) and the contributions to this change due to average individual change (light blue bars) and compositional change due to selective disappearance (orange bars) and new appearance of individuals in the study (yellow bars). The number of individuals considered in the three change intervals is 2115, 931 and 385 respectively. The error bars represent 95% confidence intervals. 

6.5. Change in MMSE score for the age group 100 at the population level (dark blue bars) and the contributions to this change due to average individual change (light blue bars) and compositional change due to selective disappearance (orange bars) and new appearance of individuals in the study (yellow bars). The number of individuals considered in the three change intervals is 903, 463 and 203 respectively. The error bars represent 95% confidence intervals.

A.1. Decomposition of changes in the mean number of doctor visits from 1996 to 1998 for German women aged 20-60 in full-time employment. The bars represent the population-level change (dark blue), the average individual change (light blue), the change due to dropout (orange) and the change due to new recruits (yellow). The error bars represent 95% confidence intervals.

A.2. Decomposition of changes in the proportion experiencing ‘bad health’ from 1996 to 1998 for German women aged 20-60 in full-time employment. The bars represent the population-level change (dark blue), the average individual change (light blue), the change due to dropout (orange) and the change due to new recruits (yellow). The error bars represent 95% confidence intervals.

A.3. Decomposition of changes in schizophrenia severity score from one observation time to the next in all data (panel 1), in the non-informatively censored group (panel 2) and in the informatively censored group (panel 3). The label on the x-axis is the follow-up week to which the change was measured. The bars represent the population-level change (dark blue), the average individual change (light blue) and the change due to dropout (orange). The error bars represent 95% confidence intervals.
A.4. Decomposition of changes in proportion of children classified as obese from one 2-year age-group to the next. The label on the x-axis is the age at the beginning of the age-group to which the change was measured. The bars represent the population-level change (dark blue), the average individual change (light blue), the change due to dropout (orange) and the change due to new recruits (yellow). The error bars represent 95% confidence intervals.

A.5. Decomposition of changes in the rate of employee absenteeism in a German company from one 6-month period to the next. The bars represent the population-level change (dark blue), the average individual change (light blue), the change due to dropout (orange) and the change due to new recruits (yellow).
List of Tables

2.1. Parameter values for the best model for females and for males . . . . . 50
2.2. Other species than the common tern for which reproductive improve-
ment has been found . . . . . . . . . . . . . . . . . . . . . . . . . . . 54
2.3. Other species than the common tern for which reproductive senes-
cence has been found . . . . . . . . . . . . . . . . . . . . . . . . . . . 58
3.1. Variables, symbols and their definition . . . . . . . . . . . . . . . . 79
3.2. Estimated parameters and their definition . . . . . . . . . . . . . . . 82
3.3. Parameter values for the model with individual effects . . . . . . . . 83
3.4. Parameter values for the model without individual effects . . . . . . 83
5.1. Decomposition of population change in relative reproduction from
event $x$ to $x + 1$ in percent . . . . . . . . . . . . . . . . . . . . . . 132
5.2. AIC for the linear mixed models to test which characteristic explains
the change in breeding performance best . . . . . . . . . . . . . . . 133
A.1. Doctor visits observation patterns . . . . . . . . . . . . . . . . . . . 167
A.2. Schizophrenia Symptoms Observation Patterns . . . . . . . . . . . 169
A.3. Childhood obesity observation patterns . . . . . . . . . . . . . . . 171
CHAPTER 1

Introduction
The study of age-specific patterns of performance is of great importance in ecology and evolutionary biology (Martin 1995; Clutton-Brock and Sheldon 2010). I investigate the mechanisms of ageing by studying changes in performance with age. The main focus of this thesis lies in examining changes in reproductive performance of a long-lived seabird species, the common tern (*Sterna hirundo*). In this introduction I provide the broad background to the topic. It is divided into five parts: First, I give an overview of the evolutionary theories of ageing. Second, I give a short summary of the evidence from the wild for the shape of age-specific reproduction in birds and mammals. Third, I describe the challenges of studying age-specific performance. Next, I introduce the datasets that are used in this thesis. Finally, I provide an overview of the thesis structure.

### 1.1. Theoretical Background

#### 1.1.1. Classic Evolutionary Theory of Ageing

Senescence is defined as declining physiological functioning with age that negatively affects the ability to survive and/or to reproduce. In the traditional understanding, senescence begins with age of maturity, and is considered unavoidable (Hamilton 1966).

There are three classic theories of ageing: the ‘mutation accumulation theory’ (Medawar 1952), the theory of ‘antagonistic pleiotropy’ (Williams 1957) and the ‘disposable soma theory’ (Kirkwood 1977). All three theories are based on the notion of decreasing selection pressure with age after reproductive maturity. Only a fraction of individuals survive to older ages and only a fraction of their fertility remains, even if there is only extrinsic mortality. This results in a decreasing net effect of a mutation with age at the population-level (Medawar 1952; Charlesworth 1973).

According to the ‘mutation accumulation theory’ of Medawar (1952), the progressive weakening in the force of selection with increasing age is eventually too weak to act against deleterious mutations that become effective at old age. Due to the strong selection pressure on early-acting mutations it is difficult for them to accumulate in the gene pool, while it is easier for late-acting mutations to escape the selection pressure. This leads to an accumulation of these mutations with late-acting deleterious effects over the generations and results in an increase of disadvantageous effects over age, which is reflected in a decreasing age-specific survival probability and fertility. The assumption behind this theory is that there are genes with age-specific actions.
The theory of ‘antagonistic pleiotropy’ predicts that pleiotropic genes which have positive fitness effects early in life are favoured by selection even if they have negative fitness effects late in life (Williams 1957). The contribution to fitness is a combination of the size of an effect and the probability of survival to the age where it becomes effective. Assuming a decreasing selection pressure, a positive effect in early life will have a stronger weight than a negative effect later in life. Due to this life history trade-off a deleterious late effect is outweighed by a favourable early effect and there is active selection for such genes. This leads again to a decreasing age-specific survival probability and decreasing fertility with age after reproductive maturity.

The underlying idea of the ‘disposable soma theory’ developed by Kirkwood (1977) is that there is a trade-off between the maintenance of the soma and the protection of the germ line. It is costly to maintain an organism. Therefore, the optimal allocation of resources between somatic maintenance and reproduction might lead to a strategy in which it is favourable to invest any spare resource into the germ cells rather than into repair of the soma, even though this eventually results in an accumulation of damage in the soma. Assuming a decreasing selection pressure with age, the pressure eventually becomes too weak for somatic maintenance to remain economical, which leads to senescence.

Hamilton (1966) derived a general measure of the force of selection to act against age-specific deleterious mutations in a theoretical model, assuming that these either reduce fertility or increase mortality. His reasoning is that the force of selection against these mutations decreases with age. Therefore, he concludes that ‘senescence is an inevitable outcome of evolution’ starting from reproductive maturity (Hamilton 1966). Williams (1957) further underlined that ‘the time of reproductive maturation should mark the onset of senescence’ for species that reach their final size latest at maturation.

Even though the classic evolutionary theories of ageing are generally discussed in terms of mortality, the same arguments should apply to reproductive performance (Hamilton 1966). According to the equations of Hamilton (1966) and Emlen (1970) the force of selection on age-specific mutations acting on fertility decreases with age. This result implies that not only mortality but also fertility should decrease over the adult lifespan (Partridge 1989; Baudisch 2005).

In ageing research, the classic view that assumes a declining selection pressure with age, has become part of biological dogma and is still widely accepted (Flatt and Promislow 2007). However, alternative view points do exist.
1.1.2. Recent Models of the Evolutionary Theory of Ageing and Life History Theory Models

Vaupel et al. (2004) stated that there are several deficiencies in the argument of Hamilton (1966), and his postulate of inevitable senescence. In particular, parameterisations for mutational effect sizes that differ from those chosen by Hamilton (1966), – for example, a proportional rather than an additive effect of mutations – leads to results where the force of selection does not necessarily need to decline with age (Baudisch 2005). Baudisch (2008) shows that the force of selection can actually increase under certain circumstances. This finding suggests that relaxation of key assumptions underpinning the classic evolutionary theory of ageing can substantially alter conclusions.

According to the vitality-based optimisation models of Baudisch (2008), it is possible to distinguish different optimal strategies that can be divided into two groups: those with senescence and those with sustenance. Vitality is defined in these models as ‘an individual’s size adjusted for the functioning of body cells’ (Baudisch 2008). An increasing vitality raises reproductive output and decreases mortality, with senescence occurring when damage exceeds repair. In sustenance strategies, repair equals or exceeds damage. When repair exceeds damage, performance increases. In theory, whenever maintenance costs are low a sustenance strategy is more advantageous than a senescent strategy (Baudisch 2008). The models of Baudisch (2008) are based on the idea of an optimal allocation of resources between growth, survival and reproduction, where optimal means a maximisation of lifetime reproductive success as a measure of evolutionary fitness. In contrast to the classic theories these models are not based on a decreasing selection pressure with age.

Starting with the work of Gadgil and Bossert (1970), several models have been developed to estimate optimal life history strategies by optimal allocation of resources among maintenance, growth and reproduction in the field of life history theory. In indeterminate growers, fecundity is typically predicted to increase with age (Kitahara et al. 1987; Kozlowski 1991; Reznick et al. 2002). Birds, however, reach their final size before starting to fly, which is long before first reproduction (Charnov 1993). Only a few life history theory models address optimal reproductive performance patterns rather than reproductive effort for determinate growers (Pianka and Parker 1975; Cichoń 2001; Kindlmann et al. 2001). They all predict a triangular pattern in reproductive performance over age: Pianka and Parker (1975) show in a graphical model that current reproduction is predicted to first increase and then decrease over age for iteroparous organisms when total lifetime reproductive success is
maximised and a trade-off between current fertility and residual reproductive value is assumed. The energy-partitioning model of Kindlmann et al. (2001) distinguishes between gonads and soma with ageing of the soma. It predicts first increasing and then declining fertility over age. The dynamic programming model of Cichoń (2001) allocates resources to growth, reproduction and repair with no overlap of growth and reproduction. It shows that high extrinsic mortality leads to an optimal strategy with high reproductive performance at the age of maturity and a continuous decline thereafter. An early increase in reproductive performance and a decrease at older ages is predicted if extrinsic mortality is low and is therefore the likely pattern for long-lived organisms (Cichoń 2001).

1.2. Evidence From the Wild

For species with indeterminate growth, where fertility is strongly connected to size, improvement in age-specific fertility over very long periods of life has been observed. The list given by Vaupel et al. (2004) includes examples from plants, algae, corals and molluscs.

The typically reported pattern of reproductive performance in birds and long-lived mammals is an initial increase with age during early adult years and a levelling off or decline at old ages (summaries in Clutton-Brock 1988; Newton 1989; Wooller et al. 1992; Forslund and Pärt 1995). Although difficult to measure due to the decreasing number of individuals alive with increasing age, reproductive senescence seems to be common in the wild and has been detected in many different species (Jones et al. 2008; Nussey et al. 2008). Table 2.3 in chapter 2 provides an overview of bird and mammal species for which senescence has been found. Please refer to table 2.2 for a list of species for which improvement in reproductive performance with increasing age has been found. Reproductive performance has been measured, for example, by the probability of reproducing (Cam et al. 2002; Wintrebert et al. 2005; Reed et al. 2008; Berman et al. 2009; Blas et al. 2009; Desprez et al. 2011), clutch size, the probability of successful breeding (Reid et al. 2003; Rattiste 2004; Weladji et al. 2006; Angelier et al. 2007c, b; Reed et al. 2008; Aubry et al. 2009; Berman et al. 2009), the probability of egg or chick survival (van de Pol et al. 2006), the number of fledglings, juveniles or recruits (van de Pol et al. 2006; McCleery et al. 2008; Blas et al. 2009; Bouwhuis et al. 2009; Brown and Roth 2009; Steenhof and Heath 2009; Knape et al. 2011) or proxies like nest initiation or laying date (González-Solis et al. 2004; van de Pol et al. 2006; McCleery et al. 2008; Brown and Roth 2009; Steenhof and Heath 2009; Auld and Charmantier 2011; Kim et al. 2011), egg volume
(González-Solís et al. 2004; Altwegg et al. 2007) or individual weight (Limmer and Becker 2007).

1.3. Challenges Addressed in this Thesis

Several aspects complicate the analysis of age-specific patterns. An important aspect of this thesis is to apply appropriate existing methods and to develop new models and methods to address several of these challenges.

1.3.1. Appropriate Model for Reproductive Data of Long-Lived Vertebrates

Changes in performance with age are best measured in longitudinal studies (Forstmeier 2002; Nussey et al. 2008). Individuals are measured repeatedly in this type of study and therefore the analysis needs to account for the non-independence of measurements within each individual. A random individual intercept can account for this non-independence (Fitzmaurice et al. 2004) and mixed models are now widely used in ecology and evolutionary biology (e.g. Reid et al. 2003; Nussey et al. 2006; Balbontín et al. 2007; Descamps et al. 2008; McCleery et al. 2008; Bouwhuis et al. 2009; Brown and Roth 2009; Nisbet and Dann 2009).

Reproductive performance measured by the number of fledglings is a count, and is usually analysed with a Poisson regression with individual random effects (van de Pol et al. 2006; McCleery et al. 2008; Blas et al. 2009; Bouwhuis et al. 2009; Brown and Roth 2009; Steenhof and Heath 2009; Knapke et al. 2011). This is the model typically recommended in standard statistical textbooks for biologists (Crawley 2007). Many long-lived vertebrates have a low maximum number of offspring per breeding season. The common terns in the studied breeding colony have a maximum of three fledglings per breeding attempt. Therefore, the Poisson distribution describes the data inaccurately and will predict, although with low probability, numbers that do not occur in reality (Kendall and Wittmann 2010; McDonald and White 2010). A limitation of the Poisson distribution is further that a single parameter determines the mean and the variance. Real data are, however, often overdispersed with a higher variance than the Poisson model allows.

Reproductive performance is a mixture of two process, the probability of breeding and the number of offspring produced. Consequently, zero fledglings can be a result of either non-breeding or breeding failure. A Poisson model cannot distinguish between these two routes to zero.
A model capturing both routes to zero fledglings, which is appropriate for counts with a low maximum possible number, measured repeatedly within individuals, should be applied to get a proper picture of age-specific reproductive performance in this species.

### 1.3.2. Pair Bond Length

Many bird species are monogamous (Lack 1968), which is also the case for the common tern. Consequently, two partners breed together and the age pattern of fertility could potentially be confounded by an effect of pair bond length (Fowler 1995). Different hypotheses have been proposed concerning why it might be advantageous to divorce or to be faithful to a known partner (see summary in section 3.2 of chapter 3). Most earlier studies have dealt with the immediate consequence for reproductive performance of divorce (Coulson 1966; Williams and Rodwell 1992; Ens et al. 1993; Dhondt and Adriaensen 1994; Briëd and Jouventin 1999; González-Solís et al. 1999a; Streif and Rasa 2001; Blondel et al. 2000; Saino et al. 2002; Moody et al. 2005; Pampus et al. 2005; Jeschke et al. 2007; Ludwig and Becker 2007; Hatch and Westneat 2008; Ismar et al. 2010). The actual pair bond length as an increasing function has only rarely been studied (Ollason and Dunnet 1978; Bradley et al. 1995; Briëd and Jouventin 1999; Pyle et al. 2001; van de Pol et al. 2006; Balbontín et al. 2007; Naves et al. 2007; Nisbet and Dann 2009).

The common tern is a monogamous species in which mate retention is high, but mate change still occurs (González-Solís et al. 1999a, c; Becker et al. 2001). While the recapture rate of breeding individuals marked with a transponder is very high at the study site (Szostek and Becker 2011), many partners are not marked and therefore are not individually identifiable. These are mostly immigrants. A model is needed that can analyse both partners simultaneously as well as their pair bond length and is able to deal with the missing information.

### 1.3.3. Heterogeneity

There are three different possible explanations for increasing reproductive success with age:

- Compositional change: Disappearance of frail individuals due to selective mortality (Vaupel et al. 1979; Vaupel and Yashin 1985; Cam and Monnat 2000) or new appearance of individuals in the dataset (van de Pol and Verhulst 2006). This new appearance may be due to immigration or postponed recruitment.
It might also be a sampling issue due to a lack of data before the onset of the study.

- Individual improvement: Improvements in experience, foraging ability or dominance status which lead to a more efficient use of resources (Weimerskirch 1992; Forslund and Pärt 1995; Newton and Rothery 1997; Wendeln 1997; Weladji et al. 2006; Angelier et al. 2007c; Limmer and Becker 2007). Physiological changes might also explain individual improvement (Heidinger et al. 2006).

- Increasing reproductive effort: Optimising the reproductive effort, which means it pays better to reproduce later given the same amount of resources invested (Forslund and Pärt 1995).

For long-lived species, like the common tern, it is unlikely that an increasing reproductive effort due to low probability of future reproduction is responsible for increasing fertility early in life (Forslund and Pärt 1995). Therefore, the two processes that need to be distinguished are individual improvement and compositional change.

Separating these processes is important because heterogeneity within a population can produce dynamics at the population level that are different from the dynamics within homogeneous subgroups or at the individual level. Frail individuals suffer a higher death rate and tend to die first, leaving the healthier individuals with a lower death rate. This selective disappearance leads to a compositional change in the population. In addition, selective appearance of individuals with certain characteristics in the study can influence the age-specific pattern at the population level again due to a compositional change in the population (van de Pol and Verhulst 2006). Therefore conclusions about individual age-trajectories often cannot be drawn directly from observations made on the aggregate population (Vaupel et al. 1979; Vaupel and Yashin 1985; Cam and Monnat 2000).

After finding a relationship between age and reproduction at the population level, some researchers conducted further analyses to capture individual-level change (Weimerskirch 1992; Laaksonen et al. 2002; Reid et al. 2003; see section 5.9.3 in chapter 5 for details). Recently, several methods have been introduced to correct for compositional change in the analysis of age-specific patterns (McNamara and Houston 1996; Cam et al. 2002; Reid et al. 2003; Rattiste 2004; Barbraud and Weimerskirch 2005; Beauplet et al. 2006; Nussey et al. 2006, 2009; van de Pol and Verhulst 2006; Weladji et al. 2006, 2010; Balbontín et al. 2007; McCleery et al. 2008; Aubry et al. 2009; Bouwhuis et al. 2009; Brown and Roth 2009; Moyes et al. 2009; see section 5.9.3 in chapter 5 for details). The most widely used approach is a mixed
model with individual random intercepts and some fixed effect as an approximation of the quality of an individual (Rattiste 2004; Nussey et al. 2006, 2009; van de Pol and Verhulst 2006; Weladji et al. 2006, 2010; Balbontín et al. 2007; McCleery et al. 2008; Aubry et al. 2009; Bouwhuis et al. 2009; Brown and Roth 2009; Moyes et al. 2009). Moyes et al. (2009) showed that the choice of the fixed effect can influence the results.

Several authors have speculated about the importance of the different factors creating the observed population-level fertility pattern (Reid et al. 2003; González-Solís et al. 2004; Rattiste 2004; Bowen et al. 2006; Weladji et al. 2006; Angelier et al. 2007c; Balbontín et al. 2007; McCleery et al. 2008; Nevoux et al. 2008; Blas et al. 2009; Bouwhuis et al. 2009; Limmer and Becker 2010). It is likely that several factors act jointly (Weimerskirch 1992; Forslund and Pärt 1995). However, none of the approaches existing so far is able to quantify the relative contributions of average within individual changes, versus changes due to selective appearance, versus changes due to selective disappearance, to the change in population patterns of fertility over age. This gap should be closed in order to get a better understanding of the mechanisms forming the observed age-specific pattern of performance.

1.4. Datasets

1.4.1. Common Tern

Birds generally age more slowly than mammals and their relatively long lifespan makes them good model species to study age-specific processes (Holmes and Austad 1995). The main focus of this thesis lies in the age-specific patterns of reproduction of the common tern (see figure 1.1). This is a long-lived seabird species for which a very detailed individual-based longitudinal dataset is available. These data come from a colony in Wilhelmshaven on the German North Sea coast. Subcutaneously implanted transponders and an automatic detection system ensure a recapture probability close to one (Becker et al. 2008b; Szostek and Becker 2011). In addition, many individual covariates (e.g. weight, sex, partner) are recorded (Becker et al. 2001, 2008b). Improvement in performance at early ages has been previously shown in the following measurements: arrival (Ezard et al. 2007; Becker et al. 2008a) and laying date (Nisbet et al. 2002; González-Solís et al. 2004; Ezard et al. 2007), egg volume (González-Solís et al. 2004), clutch size (Nisbet et al. 2002; Limmer and Becker 2010), the probability of hatching (Limmer and Becker 2010) and of fledging success (Limmer and Becker 2010) and the number of fledglings (Nisbet et al. 2002; Ezard
et al. 2007; Limmer and Becker 2010). Some of the measures were year-standardised (Ezard et al. 2007; Limmer and Becker 2010). A senescent decline at old age could not always be found for the common tern. González-Solís et al. (2004) reported a decline in egg volume starting from age 13 but did not find a delay in laying date. In the study by Nisbet et al. (2002) the laying date was earlier and the clutch size and number of fledglings was larger in common terns of age 18 and above than in birds between ages 6-9. Jones et al. (2008) were unable to detect senescence in recruitment. Ezard et al. (2007) reported a tendency for a delay in year-standardised laying date at old age but could not find a delay in year-standardised arrival date or decline in number of fledglings.

1.4.2. Danish 1905-Cohort and the Chinese Longitudinal Health and Longevity Survey

In chapter 6, two surveys are analysed to study change in functioning in elderly people. The first dataset consists of information for all Danes born in 1905, who were
still alive and living in Denmark in 1998. This longitudinal multi-assessment survey had 4 waves from 1998 to 2005. Information on the ability to perform activities of daily living independently and Mini-Mental State Examination (MMSE) scores as a measure of mental ability were gathered (Christensen et al. 2008). Independence as a combination of mental and physical ability at old age has been studied previously for this dataset (Christensen et al. 2008) and the aim is to reanalyse the published information with a new method.

In the Chinese Longitudinal Health and Longevity Survey (CLHLS) people of age 65 and above were interviewed in 1998 and in 3 subsequent waves in 2000, 2002 and 2005. Different information on health status and the quality of life were collected Zeng et al. (2002). In this thesis, only the information for the oldest old (i.e. people of at least age 80) are analysed. Scores from an adapted version of the MMSE questionnaire are used to study change in mental ability.

Both datasets were chosen because of their large sample size for old ages. The Danish study looked at an entire cohort that reached the age of 92/93. The CLHLS conducted approximately 59,000 interviews in the four waves between 1998 and 2005 of people at least 65 years old. It is the study with the largest number of centenarians worldwide (Koenig 2001).

### 1.4.3. Epidemiological and Economic Datasets

In the appendix A four different epidemiological and economic datasets are analysed to illustrate the broad usefulness of a decomposition method that is newly developed in this thesis. The datasets are on 1) the number of doctor visits before and after the German medical funding reform (the dataset `drvisits` from the R-package `sabreR`), 2) the effect of drug treatments on schizophrenia (http://www.lancs.ac.uk/staff/diggle/APTS-data-sets/PANSS_short_data.txt), 3) childhood obesity (Woolson and Clarke 1984; http://biosun1.harvard.edu/~fitzmaur/ala/obesity.txt) and 4) unemployment and absenteeism in a German company (Kauermann and Ortlieb 2004; http://www.blackwellpublishing.com/rss/Volumes/Cv53p2.htm). All datasets are publicly available.

### 1.5. This Thesis

In this thesis I address the challenges outlined above (section 1.3), that complicate the analysis of age-specific performance, to characterise patterns and get insights into why the patterns arise. The methods involve statistical modelling and establishing mathematical relationships.
In chapter 2 I apply a proportional odds model with individual random effects for repeated ordinal responses (Tutz and Hennevogl 1996; Agresti 2002) to the number of fledglings produced by common terns. Counts with a small number of possible values due to a low maximum can be considered as an ordinal response (Thomson et al. 1998; McDonald and White 2010). In this model the reproductive performance is separated into the categories not breeding, breeding without success, breeding and producing one fledgling up to breeding and producing three fledglings. The proportional odds model is thus appropriate to combine the processes of the probability of breeding and producing a certain number of offspring. I test for age and environmental effects on the reproductive performance. The proportional odds model is compared to the commonly used approach – a Poisson regression with random effect.

In chapter 3 I analyse whether the pair bond length has an effect on the number of fledglings produced by common terns. A hierarchical Bayesian model is developed that includes all steps from being alive, being present in the colony, deciding to breed, choosing a partner, producing a certain number of fledglings to the decision of breeding repeatedly. This approach handles unknowns by imputing the missing values. I test for an effect on the produced number of fledglings of the age of both partners simultaneously and the length of time they have bred together. I further study the mechanism of partner choice among other parameters of biological importance.

While the proportional odds model in chapter 2 and the hierarchical Bayesian model in chapter 3 account for the non-independence of the repeated measures within individuals they do not explicitly account for compositional change within the population. I develop a new retrospective decomposition method that can quantify the contributions of average within-individual change and compositional change to the observed population-level pattern. In chapter 4 I show how this decomposition can be derived from the Price equation (Price 1970). In this chapter only average individual change and change due to selective disappearance are separated. The decomposition is described in detail in chapter 5 and an extension including a term for selective appearance of individuals is introduced. Further, the decomposition method is applied in this chapter to reproductive performance in the common tern to disentangle whether the change with age reflects genuine changes within individuals or is an effect of a compositional change.

In chapter 6 I apply the decomposition method to two different datasets to study change in age-specific mental and physical functioning in oldest old people. In particular I analyse the change in the ability to live independently in the Danish...
1905-cohort and the change in cognitive functioning for Chinese of at least age 80 participating in the CLHLS. The appendix A further illustrates the wide range of applications of the decomposition. It is applied to four different, publicly available, epidemiological and economic datasets.

Chapter 7 gives an overall discussion drawn across this entire thesis. Additionally, it states future research directions and ends with a comprehensive conclusion.
Introduction
CHAPTER 2

Interactive Effects of Age and Environmental Conditions on Reproductive Performance in a Long-Lived Bird

This chapter will be submitted for publication under the authorship of Maren Rebke, Tim Coulson, Peter H. Becker and Jutta Gampe.

Co-author contributions: Peter H. Becker collected and provided the data. Tim Coulson conceived the basic idea of analysing the data with a generalised linear mixed model. Jutta Gampe pointed out that a proportional odds model with random effects is a more appropriate model than the Poisson regression with random effects. She further gave advice on the statistical analysis and rewrote a few paragraphs of the way the model is formally specified in the methods part of the chapter. All co-authors commented on the manuscript.
2.1. Summary

In this study we aim to analyse changes in reproductive performance with age in a long-lived bird species – the common tern (*Sterna hirundo*). Two processes determine reproductive performance: the probability of breeding and the number of fledglings produced. Like in many birds and mammals the maximum number of offspring per breeding season is low in the common tern. To analyse this type of data we use a proportional odds model with random individual effects. We fit a linear spline function of the number of fledglings over age with knots at different ages at which the slope of the function changes. We further allow for interaction with the environmental quality of the year. This model has been neglected so far in ecology and evolutionary biology and is compared to the commonly used Poisson model. Our results show sex and environmental specific differences in the age-specific reproductive pattern. Improvement in reproductive performance was evident in both sexes. Senescence, in contrast, was found in females for all years but only in good years for males. Especially individuals of intermediate age benefit disproportionately from breeding in a good year. The Poisson model, which is based on more restrictive assumptions, leads to considerably different results: The linear spline functions change for both sexes with senescence being predicted in females for good years only but for all years in males. It further strongly underestimates the variance of the random individual effects. The Poisson model thus should not be used for the analysis of repeated ordinal data with a low upper bound.

2.2. Introduction

The shape of life history patterns is frequently studied in ecology and evolution to gain insight into the mechanisms of ageing. Studying age-specific patterns of performance is important to understand life histories but also for conservation or management of species (Martin 1995; Clutton-Brock and Sheldon 2010).

We use a long-term dataset on a marked population of a long-lived seabird, the common tern (*Sterna hirundo*), to study whether and how the number of fledglings, which is used as a measure of reproductive performance, changes with age. Reproductive performance in birds and mammals is usually hump-shaped with an increase at early ages and a decline at old ages due to senescence (summaries in Clutton-Brock 1988; Newton 1989; Wooller et al. 1992; Forslund and Pärt 1995). The pattern of reproductive success in monogamous species of these animal groups is typically similar in males and females as compared to the large sex differences in polygy-
nous vertebrates (Clutton-Brock and Isvaran 2007). The common tern is a socially monogamous bird in which extra-pair paternity is rare (González-Solís et al. 2001). We will analyse both sexes separately to check if the result from the comparative study of Clutton-Brock and Isvaran (2007) is also true for the common tern or if we instead find large differences between males and females in their reproductive patterns over age.

Environmental variation causes fluctuations in reproductive success between years in the studied common tern colony (Becker 1998; Dänhardt and Becker 2011). We would firstly like to test if in our study the number of fledglings for an individual depends on the quality of the year. Secondly, if there is an effect of the quality of the year does it influence the overall level of the age-specific reproductive pattern and thus all ages equally or are certain ages affected more strongly than others?

2.2.1. Models to Study Age-Specific Reproductive Performance

Many longitudinal studies investigating reproductive performance over age analyse other measures of reproductive performance than the number of fledglings. Commonly used measures are laying date or nest initiation date (González-Solís et al. 2004; van de Pol et al. 2006; McCleery et al. 2008; Brown and Roth 2009; Steenhof and Heath 2009; Auld and Charmantier 2011; Kim et al. 2011), egg volume (González-Solís et al. 2004; Altwegg et al. 2007), breeding probability (Cam et al. 2002; Wintrebert et al. 2005; Reed et al. 2008; Berman et al. 2009; Blas et al. 2009; Desprez et al. 2011), the probability of successful breeding either in general or conditional on having layed eggs (Reid et al. 2003; Rattiste 2004; Weladji et al. 2006; Angelier et al. 2007c, b; Reed et al. 2008; Aubry et al. 2009; Berman et al. 2009) and the probability that eggs or chicks survive (van de Pol et al. 2006). These measures were mostly analysed by applying a linear mixed model or a binomial regression with random effects respectively – both methods are widely used and readily implemented into standard statistical software. Another common strategy is to standardise measures of reproductive performance by subtracting the average of that performance measure for the population in a particular breeding year and occasionally dividing additionally by the standard deviation of the measure for the population in that year. This standardisation has been done for count data, like the number of offspring, but also for other types of performance measures (Reid et al. 2003; Nussey et al. 2006; Balbontín et al. 2007; Ezard et al. 2007; Descamps et al. 2008; Nisbet and Dann 2009; Limmer and Becker 2010; Rebke et al. 2010). The standardised measure is then typically analysed with a linear mixed model (Reid et al. 2003; Nussey et al. 2006; Balbontín et al. 2007; Descamps et al. 2008; Nisbet
and Dann 2009). If count data such as clutch size, number of fledglings or juveniles or number of recruits are analysed the usual strategy is to assume that the counts follow a Poisson distribution whose parameter depends on covariates (van de Pol et al. 2006; McCleery et al. 2008; Blas et al. 2009; Bouwhuis et al. 2009; Brown and Roth 2009; Steenhof and Heath 2009; Knape et al. 2011) – the typically recommended model in statistical textbooks for biologists for count data (Crawley 2007). One of the limitations of the Poisson assumption is that a single parameter determines both the mean and the variance. So called overdispersion, that is, a variance which is higher than allowed by the Poisson model, often is observed in real data (underdispersion is less frequent). Occasionally these measurements were assumed to be normally distributed and analysed with a linear mixed model (Altwegg et al. 2007; Descamps et al. 2008; Auld and Charmantier 2011).

2.2.2. Appropriate Model

Ordinal Data Type

While it is obvious that certain quality measures such as the fat score for migrating birds are of an ordinal data type it is less intuitive why reproductive data might be also of an ordinal nature. In many long-lived organisms there is a low upper bound on the number of offspring per breeding event. In particular, in many seabird species this limit is only a few fledglings. Therefore, a Poisson error structure is inappropriate, because the model would predict, although with low probability, more offspring than the biological upper bound permits (Kendall and Wittmann 2010; McDonald and White 2010). When a clear maximum occurs at a small number, the number of offspring could instead be considered as a multinomial response and due to the natural ordering even more specifically as an ordinal response (Thomson et al. 1998; McDonald and White 2010). There are a few studies that have followed this advice (Paradis et al. 2000; Siriwardena et al. 2001; Söderström 2001; Ringsby et al. 2002; Qvanström et al. 2004; Delhey et al. 2007). All of these did not have repeated measures in their analysis or did not account for the longitudinal nature of the data. An alternative approach that fits reproductive data of vertebrate species well is the use of the generalised Poisson distribution, which has an upper bound (Kendall and Wittmann 2010).

Within a population not all individuals breed at every possible occasion. Some individuals might start reproducing at a later age than others (Ringsby et al. 2002) or might skip a year of breeding (Sanz-Aguilar et al. 2011). Zero number of fledglings can consequently result from not breeding or breeding without success. These two
sources cannot be distinguished in a pure Poisson model. The excess of zeros resulting from mixing both processes is often handled by using a zero-inflated Poisson distribution or a two-part model combining a Bernoulli and a Poisson part, which both are mixture models combining a point mass at zero with a possibly truncated Poisson distribution (reviewed in Martin et al. 2005). When reproduction is considered as an ordinal response both processes can be included as two different response categories, where not breeding at all is a worse performance than breeding without success.

**Longitudinal Study**

To study life history changes over age it is advantageous to use a longitudinal dataset (Forstmeier 2002; Nussey et al. 2008). In these longitudinal studies the life history trait, such as reproductive performance, is measured repeatedly for each individual. A random individual intercept can account for both differences between individuals due to heterogeneity (Vaupel et al. 1979; Vaupel and Yashin 1985) and at the same time capture the non-independence of responses within each individual (Fitzmaurice et al. 2004). The variance of the random effects give an estimate of how much heterogeneity there is (Agresti et al. 2000).

**Proportional Odds Model With Random Effects**

Repeated ordinal responses like annual reproductive performance with a low maximum number, are usually modelled by the proportional odds model with individual random intercepts (Tutz and Hennevogl 1996; Agresti 2002). Such a model is common in biomedical studies. To our knowledge a proportional odds model for a repeated ordinal response has not been previously used in ecology or evolutionary biology.

Liu and Agresti (2005) review different options to analyse ordinal data, of which the proportional odds model is the most popular way (Agresti and Natarajan 2001; Liu and Agresti 2005). It was originally developed by McCullagh (1980). Due to computational complexity the development of an extension into a mixed model with random effects took some time (Agresti and Natarajan 2001). This extension was introduced by Tutz and Hennevogl (1996). Different ways of analysing ordinal responses with repeated measures from a longitudinal study are summarised in Agresti and Natarajan (2001). Only rarely are closed form solutions available for mixed effect models for ordinal responses and numerical algorithms are needed to maximise the approximated likelihood. Agresti and Natarajan (2001) recommend using adaptive Gauss-Hermite quadrature (Pinheiro and Bates 1995). The propor-
tional odds model with random effects is not implemented in a maximum likelihood framework into many statistical softwares. With a modification of the likelihood it has been possible for a while to estimate a proportional odds model with random effects with the procedure **PROC NLMIXED** in **SAS**. However, the implementation is not straightforward. Several papers have been published that describe how to adapt the code in order to run a proportional odds model with random effects (Agresti et al. 2000; Sheu 2002; Yang et al. 2006). A cumulative logit model with random effects can be fitted within a Bayesian framework with the **R**-function **cumlogitRE** of the package **glmmAK** developed by Komárek and Lesaffre (2008). Only recently was the package **ordinal** with the function **clmm** developed for **R** to analyse a proportional odds model with random effects with the maximum likelihood approach (Christensen 2010).

**Linear Spline**

The usual approach to study senescence is by fitting a quadratic (Cam et al. 2002; Reid et al. 2003; González-Solís et al. 2004; Wintrebert et al. 2005; Angelier et al. 2007c; Balbontín et al. 2007; Bouwhuis et al. 2009; Brown and Roth 2009; Nisbet and Dann 2009; Kim et al. 2011) or rarely cubic (McCleery et al. 2008) curve over age. This approach can be biased because the quadratic relationship found over age might be caused by the shape in early ages where the number of individuals is high (Descamps et al. 2008; McCleery et al. 2008; Nisbet and Dann 2009; Kim et al. 2011). Therefore, several authors either used generalised additive models (Mysterud et al. 2002; Bowen et al. 2006; Ezard et al. 2007; Jones et al. 2008) or a piecewise linear regression approach with different slopes before and after certain points, the so called knots (Toms and Lesperance 2003; Weladji et al. 2006, 2010; Descamps et al. 2008; Berman et al. 2009; Kim et al. 2011; Knape et al. 2011; Millon et al. 2011). As we do not want to enforce the quadratic pattern we use such a linear spline approach for our analysis. The linear spline has knots at different ages at which the slope of the function changes, with linear pieces of the function between the knots.

**2.2.3. Outline**

Using a proportional odds model with a random individual intercept and a linear spline over age we aim to examine the age-specific pattern in the number of fledglings and the impact of environmental fluctuation for males and females in the common tern. We compare our results to the commonly used generalised linear mixed model
with a Poisson error structure.

2.3. Methods

2.3.1. Study Details

For this analysis we use information on the breeding status and the number of fledglings per breeding season of 601 female common terns with a total of 3029 observations and 622 males with 3279 observations. Most individuals have only one breeding attempt per breeding season but occasionally more attempts occur (González-Solís et al. 1999b; Becker et al. 2001; Becker and Zhang 2011). In our study we consider the overall number of fledglings per year. The data were gathered in a long term study at the Banter See in Wilhelmshaven (53°27'N, 08°07'E) on the German Wadden Sea coast from 1992 to 2009. The monospecific colony is located on six artificial islands, each of the size 10.7×4.6 m and surrounded by a 60-cm high wall. They are alligned perpendicular to the lake shore 0.9 m apart.

During the 3-month long breeding season, which extends from mid-May to mid-August, each nest in the colony is checked for new eggs, hatched and fledged chicks every 2-3 days. A fledgling is defined as a chick that survived beyond 18 days and is not found dead during the remainder of the breeding season. This ongoing study is described in detail in Becker et al. (2001).

Common terns have been ringed since 1980 in other sites in Wilhelmshaven and with the establishment of the colony since 1984 at the Banter See. Since 1992, 101 breeding adults and all chicks born in the colony have been additionally marked with a subcutaneously implanted passive integrated transponder (PIT; type: TROVAN ID 100; TROVAN, Köln, Germany). The chicks are marked shortly before fledging when they are about 14-18 days old (Becker and Wendeln 1997). For the birds PIT-tagged as chicks the exact age is known. Due to the earlier start of ringing, age information is also available for many of the birds PIT-tagged as adults. There are only 23 males and 27 females in the dataset for which the exact age is not known. As most common terns start breeding when they are 3 years old (Becker et al. 2001, 2008a), we assume that they are of this age at the first breeding observation. In addition to the marked common terns there are also many unmarked birds in the colony which are typically immigrants. In our study we only consider individuals marked with a transponder and of known sex. The sex was initially determined by behavioural observations of the adult birds and since 1998 for all PIT-tagged chicks by standard molecular methods (Becker and Wink 2002).
Each transponder has an individual-specific alphanumeric code that is read by an automatic detection system within the colony. Antennae reading the signals of transponders within a distance of ≤11 cm every 5-10 seconds are permanently located on 44 elevated resting platforms. Additional antennae are placed temporarily around each nest for 1-2 days during the breeding season to identify the partners belonging to the nest. More details of this automatic detection system are given in Becker et al. (2001, 2008b). It ensures a recapture probability very close to one (Becker et al. 2008b; Szostek and Becker 2011) and we therefore do not include recapture probability into our model. As common terns show a high site fidelity after first breeding (Becker et al. 2008b), we assume that if a bird is not detected in intermittent years it is not breeding.

2.3.2. Model Details

Proportional Odds Model with Random Effects

The number of fledglings per year for the different individuals is an unbalanced longitudinal data type with repeated measures for the individuals. It is a discrete response with more than two alternatives. Due to biological constraints the number of fledglings is usually at most 3, 4 fledglings were observed only twice. Because of the low statistical power for such a low occurrence we considered these two cases as 3 fledglings for the analysis. The response is divided into the following ordered categories: not breeding, breeding and 0 fledglings, breeding and 1 fledgling, breeding and 2 fledglings, breeding and 3 fledglings.

We apply a proportional odds model for a repeated ordinal response (Tutz and Hennevogl 1996; Agresti 2002) to these data. Age is incorporated as a fixed effect, allowing different slopes for different parts of the function with defined knots. We test different models with changes in the slope at different ages. We start using a linear spline over age, with knots at age 5, 7, 9, 11 and 13. Environmental condition in three categories is also included as a fixed effect into the model. Years are divided into categories according to the average breeding success for all marked and unmarked breeding birds within the colony. Years with an average breeding success that was lower than 0.6 are classified as bad. These are the years 2002, 2004, 2005, 2006, 2007 and 2009. Medium years have a breeding success of 0.6 to 1.29, which are the years 1993, 1996, 1997, 1998, 2000, 2003 and 2008. Good years are defined by an average breeding success larger than 1.29. These are the years 1992, 1994, 1995, 1999 and 2001. We use the medium category as a reference category. The average breeding success over time for the Banter See colony is summarised in
We include random intercepts for the individuals into our model. Both sexes are analysed separately to avoid replication of data points for members of a pair. The full model thus has random effects for the individuals, all 6 age break points and an interaction of age and the year categories.

The cumulative logits of the proportional odds model are specified in the form:

$$\logit[P(Y \leq j \mid x)] = \log \frac{P(Y \leq j \mid x)}{1 - P(Y \leq j \mid x)} \quad j = 1, \ldots, J - 1,$$

where $Y$ is the response with $J$ categories, here a total of 5, and $x$ is the set of predictors with the corresponding parameters $\beta$. $P(Y \leq j \mid x)$ are the cumulative probabilities:

$$P(Y \leq j \mid x) = \pi_1(x) + \ldots + \pi_j(x) \quad j = 1, \ldots, J - 1.$$  

The model has the general form:

$$\logit[P(Y \leq j \mid x)] = \alpha_j + \beta x \quad j = 1, \ldots, J - 1,$$

where each cumulative logit has its own cut point $\alpha_j$ but a shared parameter vector $\beta$ for the fixed covariates $x$ (Agresti 2002).

If individuals are allowed an individual random intercept, the specification 2.3 changes to

$$\logit[P(Y_{ik} \leq j \mid x_{ik}, u_i)] = (\alpha_j + u_i) + \beta x_{ik} \quad j = 1, \ldots, J - 1; \quad i = 1, \ldots, N; \quad k = 1, \ldots, n_i,$$

where $Y_{ik}$ it the $k^{th}$ observation of an individual $i$, $x_{ik}$ is the corresponding vector of fixed covariates for this observation and $u_i$ is the individual-specific realisation of the random effect, which is normally distributed with mean zero and variance $\sigma_u^2$.

From the logit-specification 2.3 the probabilities $\pi_j(x)$ for the categories $j$, $j = 1, \ldots, J - 1$, given the covariates $x$ (assuming the random effect at its mean $u_i = 0$) can be expressed as

$$\pi_j(x) = \frac{\exp(\alpha_j + \beta x)}{1 + \exp(\alpha_j + \beta x)}$$

$$\pi_1(x) = \frac{\exp(\alpha_1 + \beta x)}{1 + \exp(\alpha_1 + \beta x)} - 0$$

(2.5)
and \( \pi_j(x) = 1 - \frac{e^{\alpha_j + \beta x}}{1 + e^{\alpha_j + \beta x}} \).

The linear predictor \( \beta x_{ik} \) in our application is of two components, a linear spline for the age of the individual \( i \) at occasion \( k \) and an indicator-variable for good, medium (= reference) and bad years, respectively. An interaction term between the two components allows for environment-specific slopes of the linear-spline components.

We simplify the model stepwise and compare the models at each step via a likelihood ratio test with a \( P \leq 0.05 \) level. We first test for differences between individuals by comparing the full model to a generalised linear model without the random effects. Next, we test for environmental effects by first removing the interaction of the year category and age and if necessary by subsequently removing the additive effect of the year category. Whenever the simplification does not change the fit of the model to the data, the simplified model is accepted. The age structure in the model is subsequently simplified stepwise by removing one knot at a time from the earliest to the latest knot. Again, the simplified model without the knot is accepted when the deletion does not change the model fit otherwise the knot stays in the model. From the chosen model the next knot is then removed. This procedure is repeated until the best model for the age function is found, the most simplified possible model being a linear change over age without any knots.

To fit the model to the data we use the function \texttt{clm} for the generalised linear model and \texttt{clmm} for the generalised linear mixed model of the package \texttt{ordinal} (Christensen 2010) in the statistical software \texttt{R} version 2.11.0 (R Development Core Team 2010). The function \texttt{clmm} maximises the likelihood by adaptive Gaussian quadrature and delivers the exact maximum likelihood estimates of the parameters (Christensen 2010).

**Comparison with a Poisson Model**

For comparison with the proportional odds model we run a Poisson regression with random intercepts. For this analysis we transform the response categories into the number of fledglings, where we assign zero fledglings to the first category of not breeding. The full model has the same structure as the full proportional odds model with random effects for the individuals, all 6 age knots and an interaction of age and the year categories. We start the model simplification by testing for the interaction of age and year and proceed as described for the proportional odds model. For model fitting we use the function \texttt{lmer} from the package \texttt{lme4} in \texttt{R} (Bates and Maechler 2010), with a maximisation of the likelihood by adaptive Gaussian quadrature. To
Figure 2.1.: The linear predictor for good (blue dotted line), medium (black line) and bad years (red dashed line) for females (left) and males (right).

compare the results of the Poisson model and the proportional odds model we sum the products of the estimated probabilities for each category from the proportional odds model and the number of fledglings for each category that is used for the estimation of the Poisson model for each age.

## 2.4. Results

### 2.4.1. Proportional Odds Model with Random Effects

The comparison with a generalised linear model via a likelihood ratio test reveals that the random intercepts for the individuals are required (likelihood ratio test, $P \leq 0.001$). The estimated variance of the random intercepts is similar for both sexes with $\sigma_u^2 = 1.3143^2$ for females and $\sigma_u^2 = 1.3077^2$ for males.

For both sexes the environmental effect needs to be included as an interaction with age. This means that the effect of the environment is different for different ages (figure 2.1). There is not just a shift in the intercept of the linear predictor depending on the year category. The advantage of breeding in a good year is much higher for intermediate ages (ages 5-10) than for young or old ages.

The best model for the females has a knot at age 5, age 7 and age 13. In contrast, the best model for the males has a knot only at age 5. The parameter estimates for the best models are given in table 2.1. Both sexes show a strong increase in reproductive performance until age 5 independent of the environmental conditions (figure 2.1). After age 5 the picture is different for males and females. The results for
### Table 2.1.: Parameter values for the best model for females and for males

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Females</th>
<th>Males</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>SE*</td>
</tr>
<tr>
<td><strong>Intercepts</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>5.1450</td>
<td>0.3564</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>7.4994</td>
<td>0.3780</td>
</tr>
<tr>
<td>$\alpha_3$</td>
<td>9.6674</td>
<td>0.3927</td>
</tr>
<tr>
<td>$\alpha_4$</td>
<td>11.9752</td>
<td>0.4188</td>
</tr>
<tr>
<td><strong>Slope age</strong></td>
<td>$\beta_1$</td>
<td>1.4748</td>
</tr>
<tr>
<td><strong>Slope changes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>knot age 5</td>
<td>$\beta_{2,5}$</td>
<td>-1.5476</td>
</tr>
<tr>
<td>knot age 7</td>
<td>$\beta_{2,7}$</td>
<td>0.2278</td>
</tr>
<tr>
<td>knot age 13</td>
<td>$\beta_{2,13}$</td>
<td>-0.3583</td>
</tr>
<tr>
<td><strong>Main effect environment</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bad</td>
<td>$\beta_{3,1}$</td>
<td>1.2954</td>
</tr>
<tr>
<td>good</td>
<td>$\beta_{3,2}$</td>
<td>-0.7399</td>
</tr>
<tr>
<td><strong>Interaction age * year</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bad years</td>
<td>$\beta_{4,1}$</td>
<td>-0.5360</td>
</tr>
<tr>
<td></td>
<td>$\beta_{5,1,5}$</td>
<td>0.8824</td>
</tr>
<tr>
<td></td>
<td>$\beta_{5,1,7}$</td>
<td>-0.4527</td>
</tr>
<tr>
<td></td>
<td>$\beta_{5,1,13}$</td>
<td>0.1654</td>
</tr>
<tr>
<td>good years</td>
<td>$\beta_{4,2}$</td>
<td>0.7639</td>
</tr>
<tr>
<td></td>
<td>$\beta_{5,2,5}$</td>
<td>-1.0305</td>
</tr>
<tr>
<td></td>
<td>$\beta_{5,2,7}$</td>
<td>0.1056</td>
</tr>
<tr>
<td></td>
<td>$\beta_{5,2,13}$</td>
<td>0.2631</td>
</tr>
</tbody>
</table>

* Standard Error

The females show in bad years a further increase in reproductive performance, with a shallower slope than before, until age 13. In medium and good years reproductive performance shows a short drop from age 5 to age 7. It stabilises in good years after age 7 and increases again in medium years. Independent of the quality of the year the reproductive performance declines in females after age 13. For the males the reproductive performance increases steadily over the entire life time in medium and bad years with a shallower slope after age 5 than at younger ages. In good years, however, the pattern is similar to the pattern for the females with a decline in reproductive performance starting from age 5. In accordance with the females the level of the reproductive performance in good years nevertheless stays continuously above the level for medium and bad years. While females still have a strong advantage of breeding in a good year even at very high ages, the curves for bad, medium and good years converge at high ages for the males.

Transformed into probabilities the results show that there is a decreasing probability of non-breeding with age for males and for females in good years (figure 2.2). In good years the probability of non-breeding becomes very low from age 5 onwards. For bad and medium years the probability of non-breeding increases again in females of high age. The probability of having 0 fledglings, although breeding, first increases with age up to age 5 independent of the sex and the environmental conditions considered. It is much higher in bad and medium years than in good
Figure 2.2.: Probability of being in one of the following breeding categories over age: not breeding (dark blue), breeding and 0 fledglings (light blue), breeding and 1 fledgling (red), breeding and 2 fledglings (orange), breeding and 3 fledglings (yellow). The results are given for females and males separately in bad, medium and good years. 

years. This probability stays approximately stable after age 5 in females in bad years. It first declines after age 5 and increases again after age 13 in medium years for females. For males it decreases after age five in bad and medium years. In good years the probability of having 0 fledglings despite breeding increases slightly after age 5 in both sexes. There is a tendency for an increasing probability of having 1 fledgling for both sexes. These probabilities are lower in bad years than in medium and good years. For females the probability of producing 1 fledgling declines again at old ages in bad and medium years, but not in good years. The probability of having 2 or 3 fledglings is in general low in bad and medium years and high in good years. It is almost impossible for common terns to produce 3 fledglings in a bad year. The probabilities of producing 2 and 3 fledglings first increase in all cases up to age 5. For males these probabilities keep increasing over the entire life in bad and medium years. In contrast, for females in bad and medium years these probabilities decline at older ages starting from age 13. In good years the probabilities of having 2 and 3 fledglings reach their maximum at age 5 and decline from that age onwards, but still remain constantly at high levels compared to the other two year categories.
To see if there is a viability selection effect we compare the empirical Bayes estimates of the random effects (adapted from Pinheiro and Chao 2006) of those individuals dying at young ages to those dying at old ages. The random effect estimates are here used as a measure of the reproductive quality of the individual. As the random intercepts are normally distributed with a mean of zero an estimate greater than zero implies good reproductive quality. We analyse here only those individuals breeding at least once for which we know when they disappeared from the study. These individuals show a high fidelity to their breeding site (Becker et al. 2008a; Szostek and Becker 2011) and therefore disappearance is likely equivalent to death. Of those individuals in the lower 25 quantile of the ages of disappearance 18% of the females and 21% of the males have negative random effect estimates. While 72% of the females and 73% of the males in the upper quantile have an estimate bigger than the mean of zero.

2.4.2. Comparison with a Poisson Model

When the dataset is analysed with a Poisson model with random effects the stepwise simplification procedure leads to a different favoured model than the analysis with the proportional odds model (figure 2.3). The interaction of the year category and age is also justified in the Poisson model but the linear spline functions over age are different for both sexes than the preferred age pattern in the proportional odds model. The best model for males has a change in slope after age 5 and age 13. For females the Poisson model predicts knots at age 5 and age 7 but not at age 13. The general results for males and females are reversed. According to the Poisson model females show a slight decline in reproductive performance only in good years, while males have a decreasing performance independent of the quality of the year. As in the proportional odds model, the intermediate ages benefit disproportionately from a year with high average population breeding success in both sexes. The estimated variance of the random intercepts in the Poisson model of $\sigma^2_u = 0.2864^2$ for females and of $\sigma^2_u = 0.3476^2$ for males are much smaller than the estimates in the proportional odds model.
Figure 2.3.: The average number of fledglings for good (blue dotted line), medium (black line) and bad years (red dashed line) for females (top) and males (bottom) predicted by the proportional odds (left) and the Poisson model (right) with random effects.

2.5. Discussion

2.5.1. Our Results in Comparison to Previous Findings

Age-Specific Reproductive Pattern

Our analyses show an increase in the number of fledglings for both sexes at early ages independent of environmental conditions. An improvement in early performance has been previously shown for the common tern in arrival (Becker et al. 2008a) and laying date (Nisbet et al. 2002; González-Solís et al. 2004), in egg volume (González-Solís et al. 2004), in clutch size (Nisbet et al. 2002), in the number of fledglings (Nisbet et al. 2002) and in several year-standardised measures – arrival date (Ezard et al. 2007), laying date (Ezard et al. 2007), clutch size (Limmer and Becker 2010), probability of hatching success (Limmer and Becker 2010), probabil-
Table 2.2.: Other species than the common tern for which reproductive improvement has been found

<table>
<thead>
<tr>
<th>Species</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Northern fulmar (<em>Fulmarus glacialis</em>)</td>
<td>Ollason and Dunnet (1978)</td>
</tr>
<tr>
<td>Sparrowhawk (<em>Accipiter nisus</em>)</td>
<td>Newton et al. (1981)</td>
</tr>
<tr>
<td>Southern fulmar (<em>Fulmarus glacialis</em>)</td>
<td>Weimerskirch (1990)</td>
</tr>
<tr>
<td>Antarctic fur seal (<em>Arctocephalus gazella</em>)</td>
<td>Lunn et al. (1994)</td>
</tr>
<tr>
<td>Northern wheatear (<em>Oenanthe oenanthe</em>)</td>
<td>Pärt (2001)</td>
</tr>
<tr>
<td>Cassin’s auklet (<em>Diotichomorus aleuticus</em>)</td>
<td>Pyle et al. (2001)</td>
</tr>
<tr>
<td>Tengmalm’s owl (<em>Aegolius funereus</em>)</td>
<td>Laaksonen et al. (2002)</td>
</tr>
<tr>
<td>Domestic sheep (<em>Ovis aries</em>)</td>
<td>Mysterud et al. (2002)</td>
</tr>
<tr>
<td>Red-billed chough (<em>Pyrrhocorax pyrrhocorax</em>)</td>
<td>Reid et al. (2003)</td>
</tr>
<tr>
<td>Black-legged kittiwake (<em>Rissa tridactyla</em>)</td>
<td>Cam et al. (2002); Wintrebert et al. (2005)*</td>
</tr>
<tr>
<td>Oystercatcher (<em>Haematopus ostralegus</em>)</td>
<td>van de Pol and Verhulst (2006)**</td>
</tr>
<tr>
<td>Subantarctic fur seal (<em>Arctocephalus tropicalis</em>)</td>
<td>Beauplet et al. (2006)</td>
</tr>
<tr>
<td>Grey seal (<em>Halichoerus grypus</em>)</td>
<td>Bowen et al. (2006)</td>
</tr>
<tr>
<td>Red deer (<em>Cervus elaphus</em>)</td>
<td>Russey et al. (2006)</td>
</tr>
<tr>
<td>Reindeer (<em>Rangifer tarandus</em>)</td>
<td>Wéladji et al. (2006, 2010)</td>
</tr>
<tr>
<td>Barn owl (<em>Tyto alba</em>)</td>
<td>Altwegg et al. (2007)</td>
</tr>
<tr>
<td>Black-browed albatross (<em>Thalassarche melanophris</em>)</td>
<td>Angelier et al. (2007c)</td>
</tr>
<tr>
<td>Snow petrel (<em>Pagodroma nivea</em>)</td>
<td>Angelier et al. (2007b); Berman et al. (2009)</td>
</tr>
<tr>
<td>Barn swallow (<em>Hirundo rustica</em>)</td>
<td>Balbontín et al. (2007)</td>
</tr>
<tr>
<td>Weddell seal (<em>Leptonychotes weddellii</em>)</td>
<td>Hadley et al. (2007)</td>
</tr>
<tr>
<td>Red squirrel (<em>Tamiasciurus hudsonicus</em>)</td>
<td>Descamps et al. (2008)</td>
</tr>
<tr>
<td>Mute swan (<em>Cygnus olor</em>)</td>
<td>McCleery et al. (2008)</td>
</tr>
<tr>
<td>Black kite (<em>Milvus migrans</em>)</td>
<td>Blas et al. (2009)</td>
</tr>
<tr>
<td>Great tit (<em>Parus major</em>)</td>
<td>Bouwhuis et al. (2009)</td>
</tr>
<tr>
<td>Wood thrush (<em>Hylocichla mustelina</em>)</td>
<td>Brown and Roth (2009)</td>
</tr>
<tr>
<td>Little penguin (<em>Eudyptula minor</em>)</td>
<td>Nisbet and Dann (2009)</td>
</tr>
<tr>
<td>Northern flicker (<em>Colaptes auratus</em>)</td>
<td>Wiebe and Gerstmar (2010)</td>
</tr>
<tr>
<td>Blue tit (<em>Cyanistes caeruleus oglastrae</em>)</td>
<td>Auld and Charmantier (2011)</td>
</tr>
<tr>
<td>Collared flycatcher (<em>Ficedula albicollis</em>)</td>
<td>Evans et al. (2011)</td>
</tr>
<tr>
<td>Blue-footed booby (<em>Sula nebulosa</em>)</td>
<td>Kim et al. (2011)</td>
</tr>
<tr>
<td>Silvereye (<em>Zosterops lateralis</em>)</td>
<td>Knape et al. (2011)</td>
</tr>
<tr>
<td>Tawny owl (<em>Strix aluco</em>)</td>
<td>Millon et al. (2011)</td>
</tr>
</tbody>
</table>

* Only in breeding probability  
** Only in laying date

According to the review by Forslund and Pärt (1995) there are three different hypotheses for increasing reproductive success with age: the ‘constraint hypothesis’, the ‘selection hypothesis’ and the ‘restraint hypothesis’. All three hypotheses are
frequently discussed in the literature and there are claims for support from studies in the wild for all hypotheses (‘constraint hypothesis’: Reid et al. (2003); González-Solís et al. (2004); Rattiste (2004); Bowen et al. (2006); Weladji et al. (2006); Angelier et al. (2007c); Balbontín et al. (2007); McCleery et al. (2008); Nevoux et al. (2008); Blas et al. (2009); Bouwhuis et al. (2009); Limmer and Becker (2010); Rebke et al. (2010); ‘selection hypothesis’: Reid et al. (2003); Weladji et al. (2006); McCleery et al. (2008); Nevoux et al. (2008); Blas et al. (2009); Bouwhuis et al. (2009); Rebke et al. (2010); ‘restraint hypothesis’: Blas et al. (2009)).

Generalised linear mixed models allow for an individual-specific interpretation of parameters (Agresti et al. 2000; Agresti 2002). Therefore, the presented results are evidence of improvement in reproductive success with age within the individuals. However, there is also some indication of heterogeneity in the population. We found no tendency of the birds dying very early to be of bad reproductive quality but there is a tendency for the long-lived individuals to be of good reproductive quality. This combination of individual quality differences and experience related changes is most likely the usual case (Weimerskirch 1992; Wooller et al. 1992; Forslund and Pärt 1995; Cam et al. 2002; Reid et al. 2003; Weladji et al. 2006; McCleery et al. 2008; Nevoux et al. 2008; Blas et al. 2009; Bouwhuis et al. 2009; Rebke et al. 2010), however, there are studies that did not find any selection effect (Rattiste 2004; Bowen et al. 2006; Angelier et al. 2007c; Balbontín et al. 2007; Nisbet and Dann 2009; Kim et al. 2011; Knape et al. 2011).

Although there are studies where the apparent improvement at early ages is caused by selective appearance or disappearance of individuals (egg volume in oystercatchers (Haematopus ostralegus, van de Pol and Verhulst 2006); probability of breeding success in black-legged kittiwakes (Rissa tridactyla, Aubry et al. 2009)) an earlier study of the same common tern colony showed that these two factors play only a minor role in shaping the age-specific pattern of reproductive performance in the studied species (Rebke et al. 2010).

The increase in number of fledglings with age in the common tern might be explained by a gain in breeding experience (Coulson 1966; Forslund and Larsson 1992; Pyle et al. 2001; Cichoń 2003; Rattiste 2004; Weladji et al. 2006, 2008; Angelier et al. 2007c; Blas et al. 2009; Steenhof and Heath 2009; Limmer and Becker 2010; Desprez et al. 2011). The individuals might improve their ability to raise fledglings due to enhanced foraging performance, mate cooperation and physiological adaptations (Weimerskirch 1990; Green 2002; Cichoń 2003; Limmer and Becker 2009; Bauch et al. 2010; Zimmer et al. 2011). It has been shown that common terns increase their body mass with age, which is likely an effect of improving foraging
Interactive Effects of Age and Environment on Reproduction

efficiency (Becker and Bradley 2007; Limmer and Becker 2007). The strong increase in breeding performance until age 5 that we found for both sexes in our study also corresponds to an increase in body mass until that age (Limmer and Becker 2007). Further, the ability to cope with breeding constraints such as coordination of breeding times to avoid fasting might increase with age (Angelier et al. 2007c). Less response to stress in older birds has been found for long-lived seabirds, including the common tern, and thus older parents might be less likely to react to stress with a reduction in parental care (Heidinger et al. 2006; Angelier et al. 2007b). Another possible contributing factor is an earlier arrival and initiation of egg laying in the season with increasing age which is generally a reproductive advantage (Wendeln 1997; Arnold et al. 2004; González-Solís et al. 2004; Becker et al. 2008a). It is unlikely that the restraint hypothesis, which suggests an increasing reproductive effort with age because of decreasing future reproductive perspectives, explains the increasing reproductive performance early in life for a bird with high adult age-specific survival probability (Forslund and Larsson 1992; Forslund and Pärt 1995).

Common terns typically first return to the colony as prospectors when they are at least two years old, with rare exceptions of one year old prospectors (Dittmann and Becker 2003; Becker et al. 2008b). This prospecting behaviour is also evident in our analysis via the probability of non-breeding in the first age classes. Especially in later age classes this probability is evidence for skipping behaviour, which is most pronounced in bad years and is rare in good years. Szostek and Becker (2011) found in an analysis where both sexes were combined, that skipping is more evident in old ages than in intermediate ages in the common tern. This result is in agreement with our finding that the probability of non-breeding increases again in females at old age.

Not all the prospectors necessarily recruit to the colony they visited (Dittmann et al. 2007). Therefore, disappearance before recruitment to the breeding population does not necessarily mean death. After first breeding the site fidelity of common terns is very high (Becker et al. 2008a; Szostek and Becker 2011). Most common terns start breeding at age 3. There are a few that recruit already at age 2 and a considerable proportion only at age 4. Only 11% recruit between the ages of 5 and 7 (Becker et al. 2001, 2008b; Ludwigs and Becker 2002). The slight drop in reproductive performance of females in medium and good years might be an artefact caused by the late recruits since not only age but also experience was found to be an important determinant of reproductive performance in the common tern (Limmer and Becker 2010). In addition, late recruits were shown to have a slower increase in reproductive performance than early recruits (Limmer and Becker 2010).
While in this study females show senescence independent of the quality of the year males only show a decline in reproductive performance in good years after reaching an extraordinary high performance at age 5. Previous studies on reproductive performance of the common terns do not give a concise picture. Rebke et al. (2010) provide similar results to this study. Both sexes show an increase in reproductive performance at the beginning of life. While the senescent decline in year-standardised number of fledglings was clear for female common terns from age 14 onwards, the 95% confidence intervals for males were very wide and hence the decline was inconclusive. González-Solís et al. (2004) found that egg volume stabilises after a strong increase until age 5 and starts to decline at age 13 for females. This pattern coincides with the period of strong enhancement and the starting age of senescence for females in our study. Laying date for female common terns advanced until age 8 when it reached a plateau, lacking any sign of senescence (González-Solís et al. 2004). Nisbet et al. (2002) were also unable to find senescence and instead showed an increasing reproductive performance until high ages in their cross-sectional study of one particular year. In this study, common terns of age 18 and above had an earlier laying date, a larger clutch size and a higher number of fledglings than those between ages 6-9 (Nisbet et al. 2002). Jones et al. (2008) could not find a decline in recruitment for the common tern. Ezard et al. (2007) did not detect senescence in year-standardised arrival date or number of fledglings but a tendency for senescence in year-standardised laying date starting from age 18.

What is the pattern for other species? While senescence in reproductive performance seems to be the rule in birds and mammals there are exceptions without a decline in old age. Senescence in reproductive performance has been found in various different species (table 2.3). The decline in reproductive performance at old age might be due to physical deterioration but also decreasing foraging efficiency (Catry et al. 2006; Lecomte et al. 2010; Zimmer et al. 2011).

However, not all studies were able to detect senescence in reproductive performance. Mauck et al. (2004) did not find any decline in hatching probability for Leach’s storm petrels (Oceanodroma leucorhoa) at old age. In snow petrels (Pagodroma nivea) the probability of successful reproduction stabilises after age 12 and does not decline at old age (Angelier et al. 2007b; Berman et al. 2009). In American kestrels (Falco sparverius) no sign of senescence in reproductive performance could be found (Steenhof and Heath 2009). In Southern fulmars (Fulmarus glacialoides) the decline in the proportion of eggs that produced fledglings between the age groups 18-20 and 21-26 was not significant (Weimerskirch 1990). In female barn owls (Tyto alba) the number of fledglings increases over the entire lifespan and there is no sign
Table 2.3.: Other species than the common tern for which reproductive senescence has been found

<table>
<thead>
<tr>
<th>Species</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Northern fulmar (<em>Fulmarus glacialis</em>)</td>
<td>Ollason and Dunnet (1978)**</td>
</tr>
<tr>
<td>Sparrowhawk (<em>Accipiter nisus</em>)</td>
<td>Newton et al. (1981)**</td>
</tr>
<tr>
<td>Antarctic fur seal (<em>Arctocephalus gazella</em>)</td>
<td>Lunn et al. (1994)</td>
</tr>
<tr>
<td>Black-legged kittiwake (<em>Rissa tridactyla</em>)</td>
<td>Cam et al. (2002); Rattiste (2004); Wintrebert et al. (2005)</td>
</tr>
<tr>
<td>Tengmalm’s owl (<em>Aegolius funereus</em>)</td>
<td>Laaksonen et al. (2002)*</td>
</tr>
<tr>
<td>Fallow deer (<em>Dama dama</em>)</td>
<td>McElligott et al. (2002)</td>
</tr>
<tr>
<td>Domestic sheep (<em>Ovis aries</em>)</td>
<td>Mysterud et al. (2002)</td>
</tr>
<tr>
<td>Red-billed chough (<em>Pyrrhocorax pyrrhocorax</em>)</td>
<td>Reid et al. (2003)**</td>
</tr>
<tr>
<td>Subantarctic fur seal (<em>Arctocephalus tropicalis</em>)</td>
<td>Beauplet et al. (2006)</td>
</tr>
<tr>
<td>Grey seal (<em>Halichoerus grypus</em>)</td>
<td>Bowen et al. (2006)</td>
</tr>
<tr>
<td>Red deer (<em>Cervus elaphus</em>)</td>
<td>Nussey et al. (2006, 2009)</td>
</tr>
<tr>
<td>Barn owl (<em>Tyto alba</em>)</td>
<td>Altwegg et al. (2007)*</td>
</tr>
<tr>
<td>Black-browed albatross (<em>Thalassarche melanophris</em>)</td>
<td>Angelier et al. (2007c)</td>
</tr>
<tr>
<td>Barn swallow (<em>Hirundo rustica</em>)</td>
<td>Balbontín et al. (2007)**</td>
</tr>
<tr>
<td>Red squirrel (<em>Tasmiscerus hudsonicus</em>)</td>
<td>Descamps et al. (2008)</td>
</tr>
<tr>
<td>Mute swan (<em>Cygnus olor</em>)</td>
<td>McCleery et al. (2008)</td>
</tr>
<tr>
<td>Common guillemot (<em>Uria aalge</em>)</td>
<td>Reed et al. (2008)</td>
</tr>
<tr>
<td>Snow petrel (<em>Pagodroma nivea</em>)</td>
<td>Berman et al. (2009)**</td>
</tr>
<tr>
<td>Southern fulmar (<em>Fulmarus glacialoides</em>)</td>
<td>Berman et al. (2009)</td>
</tr>
<tr>
<td>Great tit (<em>Parus major</em>)</td>
<td>Bouwhuis et al. (2009)</td>
</tr>
<tr>
<td>Wood thrush (<em>Hylocichla mustelina</em>)</td>
<td>Brown and Roth (2009)</td>
</tr>
<tr>
<td>Little penguin (<em>Eudyptula minor</em>)</td>
<td>Nisbet and Dann (2009)</td>
</tr>
<tr>
<td>Reindeer (<em>Rangifer tarandus</em>)</td>
<td>Weladji et al. (2010)</td>
</tr>
<tr>
<td>Blue tit (<em>Cyanistes caeruleus ogliastrae</em>)</td>
<td>Auld and Charmantier (2011)</td>
</tr>
<tr>
<td>Collared flycatcher (<em>Ficedula albicollis</em>)</td>
<td>Evans et al. (2011)</td>
</tr>
<tr>
<td>Blue-footed booby (<em>Sula nebouzii</em>)</td>
<td>Kim et al. (2011)</td>
</tr>
<tr>
<td>Silvereye (<em>Zosterops lateralis</em>)</td>
<td>Knape et al. (2011)</td>
</tr>
<tr>
<td>Tawny owl (<em>Strix aluco</em>)</td>
<td>Millon et al. (2011)</td>
</tr>
</tbody>
</table>

* Only in males  
** Only in females  
*** Only in breeding probability

Interestingly, for Northern fulmars (*Fulmarus glacialis*), sparrowhawks (*Accipiter nisus*), red-billed choughs (*Pyrrhocorax pyrrhocorax*) and barn swallows (*Hirundo rustica*) there was only a reproductive decline for females but not for males (Ollason and Dunnet 1978; Newton et al. 1981; Reid et al. 2003; Balbontin et al. 2007). In addition, Reed et al. (2008) found faster senescence in female than in male common guillemots (*Uria aalge*). These are results that support our findings for bad and medium quality years. Kim et al. (2011) showed for blue-footed boobies (*Sula nebouzii*) that early reproduction leads to increased senescence in males but not in females. The senescence pattern for males that is only observed in good years might therefore be a consequence of early recruitment when environmental
conditions are favourable. Such an advanced recruitment during years of low nest predation due to a high abundance of alternative prey for potential predators was found for greater snow geese (*Chen caerulescens atlantica*, Reed et al. 2003). Millon et al. (2011) showed for the tawny owl (*Strix aluco*) that reproductive performance declines much earlier in males than in females, a result that is in accordance with our findings for years of good average reproductive breeding success. Nevertheless, most studies do not find a difference in the age-specific reproductive performance over age between sexes (Rattiste 2004; Wintrebert et al. 2005; Lecomte et al. 2010; Limmer and Becker 2010; Kim et al. 2011; Knape et al. 2011) and only the inclusion of the environmental effect might reveal the variety of patterns.

The common tern is a monogamous species. Although mate changes frequently occur, mate retention is high (González-Solís et al. 1999a, c; Becker et al. 2001). Therefore, in reality the breeding performance is likely to depend on both partners and the male and female age-pattern might be influenced by the opposite sex. Modelling both sexes simultaneously in this population is challenging because of a large number of unmarked partners. While all the chicks that have been born in the colony since 1992 are marked immigrating adults are not.

In conclusion, the lack of senescence for males in medium and bad years is not an impossible result. However, the insignificance of a knot at an old age for the males might also be due to a lower number of males still alive at that age than females (35 versus 39) and a lower maximum age reached within the study (19 versus 21).

### Environmental Effects

Environmental variation can potentially have a strong impact on phenotypic traits (Ozgul et al. 2009) and explain a big part of the variation in a phenotypic trait (Altwegg et al. 2007). In the studied population there are considerable year to year differences in average reproductive success of the colony (Becker 1998, 2010; Szostek and Becker 2011). Food availability and weather conditions are key factors for the breeding success in the studied common tern colony. Predation on eggs and chicks is another factor causing low breeding success (Becker 1998). Since 2002 the years have been predominantly of low average breeding success. Food shortage due to low stocks of pelagic fish was likely the main reason for the low breeding success (Dänhardt and Becker 2011).

Some previous studies of the common tern have either standardised measures of reproductive performance (Ezard et al. 2007; Limmer and Becker 2010; Rebke et al. 2010) or included the year as a fixed effect into the analysis to account for temporal fluctuations (González-Solís et al. 2004). A shift in the level of the age-specific repro-
Interactive Effects of Age and Environment on Reproduction

Productive curve depending on the year (Rattiste 2004; Wintrebert et al. 2005; Angelier et al. 2007b; McCleery et al. 2008; Desprez et al. 2011) or the year as a random effect (Altwegg et al. 2007; Brown and Roth 2009; Auld and Charmanier 2011) have been included in several analyses of different species. An alternative approach in line with our study is to include some proxy of the quality of the year such as annual breeding success in the population (Rattiste 2004; Berman et al. 2009; Bouwhuis et al. 2009) or duration of foraging trips measuring food availability (Lunn et al. 1994). Temporal environmental covariates like breeding density (Bouwhuis et al. 2009), rainfall (Nevoux et al. 2008), oceanographic or climate conditions (Hadley et al. 2007; Nevoux et al. 2007; Lewis et al. 2009; Oro et al. 2010; Wiebe and Gerstmar 2010) and food availability (Rindorf et al. 2000; Frederiksen et al. 2005; Dänhardt and Becker 2011) have been shown to influence breeding performance.

An interaction of demographic parameters such as age and environment can potentially lead to strong fluctuations in population size (Coulson et al. 2001). Nevertheless, such an interaction has only rarely been studied. Younger individuals tend to react stronger to environmental variation than older individuals (Gaillard et al. 2000; Oro et al. 2010), a result that has been previously shown for the common tern (Ezard et al. 2006) and is also supported by our findings. Ezard et al. (2006) found that young common tern adults between ages 3 and 6 only have a high number of fledglings under good environmental conditions. In our study a good year gives an exceptionally large advantage to intermediate aged (ages 5-10) individuals, but we also find that the young ages are only likely to reproduce successfully in good years. In bad years the breeding success for all ages is so low that there is probably not much room for a further decrease. In contrast, previous studies found the opposite – stressful situations lead to a lower reproductive performance or even to non-breeding in younger adults which is not apparent in good breeding conditions (Laaksonen et al. 2002; Bunce et al. 2005). Tavecchia et al. (2005) were only able to find a cost of reproduction in years with bad winter weather conditions.

Not only ages, but also sexes might react differently to environmental variability (Oro et al. 2010). The females in our study do not only benefit disproportionately from a good breeding year at younger ages but also at old ages. In comparison, for males the reproductive performance in years of different quality converges at old age. But overall the early increase with age in reproductive performance for both sexes is still much stronger than the difference between years.

Not only environmental conditions experienced during breeding might alter the reproductive pattern over age, also environmental conditions in the year of birth were shown to have an additional effect on the age pattern of reproduction or for certain
age groups (Descamps et al. 2008; Millon et al. 2011) and the pace of senescence (Nussey et al. 2007; Reed et al. 2008). Due to complexity we did not include such a cohort effect into our model.

2.5.2. Comparison With a Poisson Model

Different authors have highlighted the importance of generalised linear ordinal regression models for ecology in the past (Thomson et al. 1998; McDonald and White 2010; Guisan and Harrell 2000). Our comparison of the two models underlines that for ordinal responses with repeated measurements and a low upper limit a Poisson regression with random effects is not appropriate. The Poisson model strongly underestimates the variance of the random intercepts and the result from the best model is considerably different from the proportional odds model. The preferred age-knots pattern is different for the females and for the males. Our comparison thus shows that previously published results might be biased.

2.5.3. Potential Future Extensions of the Model

A possible extension of our model is to implement a smooth transition instead of using the abrupt change in slope (Toms and Lesperance 2003).

In our model we assume a normal distribution of the random effects, which is a frequently made assumption (Agresti et al. 2000). Alternatively, a penalised Gaussian mixture could be used to describe the distribution of the random effects. This approach requires Markov chain Monte Carlo techniques and is much more computer intensive (Komárek and Lesaffre 2008).

As discussed in section 2.5.1, previous studies for the common tern showed that changes within the individuals are much more important than selective disappearance of individuals (Rebke et al. 2010). Nevertheless, a possible future extension of the model could be to combine the model of reproductive performance with a survival model linking both with common random effects (Ten Have et al. 2000).

2.5.4. Conclusion

This study shows that there is a strong increase in reproductive performance at early ages in both sexes. Reproductive senescence is apparent but might not occur under all circumstances. The predicted age pattern would not have been found using a Poisson model. Environmental conditions affect reproductive performance at younger ages much more strongly than at older ages.
CHAPTER 3

Better the Devil You Know: Common Terns Stay With a Familiar Partner Although Pair Duration Does Not Affect Breeding Output

This chapter will be submitted for publication under the authorship of Maren Rebke, Fernando Colchero, Peter H. Becker, and Tim Coulson.

Co-author contributions: Peter H. Becker collected and provided the data. Fernando Colchero and I developed and programmed the model together. Fernando Colchero improved and extended the description of the model. All co-authors commented on the manuscript.
3.1. Summary

In a monogamous species two partners contribute to the breeding process. How are partners chosen? Does the length of time spent with the current partner play a role for reproductive performance? Is there an additional effect of age? Do males and females show different age patterns if they are modelled simultaneously? Is there heterogeneity in reproductive performance between individuals? In this study we use a high quality dataset of the common tern (*Sterna hirundo*), a long-lived seabird, to address these questions. Despite the high quality of the data they are not perfect: it is not always obvious if unmarked partners are new or not. Further, for unmarked partners, and also for some of the marked individuals, age, sex and time of death are unknown. To handle these data limitations and model the complicated processes driving reproduction we use a hierarchical Bayesian model of the steps that lead to the number of fledglings. We include processes operating at the individual level and the pair level in the same model. We hierarchically model the bird being alive or dead, model the subsequent steps of being present in the colony, the decision to breed up to three times within a breeding season, the partner choice and ultimately the number of fledglings. Most of the processes that are modelled depend on several parameters. The results show that the age of both partners is important for reproductive performance, with similar patterns for both sexes and individual heterogeneity in reproductive performance, but pair bond length is not. Nevertheless, the terns are more likely to choose a familiar partner independent of the previous breeding outcome with that partner. This suggests that common terns have the tendency to remain with a partner that they choose at the beginning of their breeding career.

3.2. Introduction

Many bird species are monogamous and consequently two partners are involved in the breeding process (Lack 1968). What is their pattern in reproductive performance and why does it arise? Mate fidelity increases with longer adult life expectancy and thus pair bonds might be of importance especially in long-lived species (Briëd et al. 2003). Does the length of the pair bond affect reproductive output as an aspect of fitness? Is that also the mechanism for partner choice?

Pair bonds in birds are often thought to influence breeding performance. There are different hypotheses why it might be on one hand, advantageous to stay with the same parter, or on the other hand, why it might be advantageous to divorce and
choose a new partner.

Several hypotheses have been proposed that presume divorce is an adaptive strategy. Under this assumption the strategy maximising fitness of an individual should be chosen. If the expected benefits of changing the mate outweigh the potential costs the pair should divorce (Williams 1966). When partners are not well coordinated, divorce may be an adaptive advantage, which leads to pairings with more compatible mates for both partners. This theory is called the ‘incompatibility hypothesis’ (Coulson 1966). Changing the mate due to incompatibility should lead to higher reproductive output for both partners with their new mates (Choudhury 1995). Johnston and Ryder (1987) classified this hypothesis as part of the broader ‘errors of mate choice hypothesis’ according to which divorce is a consequence of a previous wrong partner choice. The ‘better option hypothesis’ predicts that one of the partners initiates the divorce due to the option of pairing with a higher quality mate and thus increasing its breeding performance. The variation in quality of individuals plays a role here. If all birds are similar the expected gain of changing the mate will be low and might not pay off (Ens et al. 1993).

According to the ‘inbreeding avoidance hypothesis’ divorce and subsequent dispersal of one of the partners might be a mechanism of inbreeding avoidance, which plays a role especially in cooperative species (Hatchwell et al. 2000).

Other theories consider divorce as non-adaptive. One of the partners may be displaced by an intruding new mate. In this case the divorce is not initiated by any of the old members of the pair. Divorce frequency should increase according to this ‘forced divorce hypothesis’ with increasing competition (Choudhury 1995). Mate change may be enforced by some external influences leading to temporary loss of the partner which is denoted as the ‘accidental loss hypothesis’ (Owen et al. 1988). The predictions of this hypothesis are that the breeding performance does not necessarily increase after the divorce and divorced pairs frequently re-mate in following years (Owen et al. 1988; Choudhury 1995). This leads directly to the next hypothesis, the ‘musical chair hypothesis’ according to which asynchronous arrival of both partners in the breeding colony might result in divorce of pairs (Dhondt and Adriaensen 1994). In this case the mate change is a side effect and not adaptive. The cause is the reproductive advantage of early breeding. Especially in species with a low survival probability the costs of waiting for a partner that might potentially be dead are very high (Rowley 1983; Choudhury 1995).

There are not only reasons for divorce but also for re-mating with the same partner. Divorce might not be an option due to the lack of available partners or territories, which is denoted by the ‘constraint hypothesis’ (Freed 1987). A repeated
mating with the same partner might increase coordination of breeding behaviour within the pair, which leads to a better breeding performance. This effect is described by the ‘mate familiarity hypothesis’ (Rowley 1983; Real 1990; Fowler 1995).

The ‘syndrome of the first-year bond’ predicts that there is an increase from the first to the second year of breeding performance but not longer. The low breeding performance in the first year is due to inexperience with that particular partner and possibly a cost of mate change. This theory views the pair formation process as important, while the ‘mate familiarity hypothesis’ is considering the long term adjustment as the process of importance (Naves et al. 2007). Pair fidelity may avoid costs e.g. in energy and time for pre-breeding behaviour such as searching for a new partner and courtship (Rowley 1983; Real 1990). Re-mating birds can thus lay their eggs earlier which often leads to higher reproductive performance (Fowler 1995). The ‘pair-bond investment hypothesis’ predicts that this already invested effort is lost when birds divorce (Fowler 1995). In addition, it excludes the risk of not having any partner (Lifjeld and Slagsvold 1988). If breeding performance increases with age but individuals cannot assess age of partners, it is advantageous to pair repeatedly with the same partner which automatically gets older. This is described in the ‘assured-age hypothesis’ (Rowley 1983; Fowler 1995). The same hypothesis might hold for increasing experience instead of age (Rowley 1983).

The common questions that studies ask are related to divorce and not pair bond length; which is what needs to be studied to test the ‘mate familiarity hypothesis’. They ask if pairs or individuals that divorce in the following year have a lower breeding success than those that stay together or if breeding failure in the previous year is predictive of divorce (Coulson 1966; Mills 1973; Williams and Rodwell 1992; Dhondt and Adriaensen 1994; González-Solís et al. 1999a; Blondel et al. 2000; Mougin et al. 2000; Becker et al. 2001; Streif and Rasa 2001; Saino et al. 2002; Saitou 2002; Moody et al. 2005; Pampus et al. 2005; Ludwigs and Becker 2007; Naves et al. 2007; Arai et al. 2009; Ismar et al. 2010). But they do not consider the accumulated previous breeding output. Divorce is rather viewed as an ad-hoc decision that is only based on the breeding output of one year. Not many studies test for an effect of pair bond length on the probability of divorce (Briëd and Jouventin 1999; van de Pol et al. 2006; Naves et al. 2007; Hatch and Westneat 2008; Nisbet and Dann 2009). We could not find any other study that considers all previous mates for the mating decision and thus analyses the overall partner choice mechanism rather than just the decision to split up with the previous partner.

The change in breeding success of only divorced or faithful and divorced individuals is usually compared in two adjacent breeding seasons (Coulson 1966; Ens et al.
Better the Devil You Know

1993; Dhondt and Adriaensen 1994; González-Solís et al. 1999a; Streif and Rasa 2001; Blondel et al. 2000; Moody et al. 2005; Pampus et al. 2005; Jeschke et al. 2007; Ludwigs and Becker 2007) or the breeding success is compared directly between individuals that divorced and those that re-mated with the previous partner to study the effect of divorce (Williams and Rodwell 1992; Ens et al. 1993; Briëd and Jouventin 1999; González-Solís et al. 1999a; Blondel et al. 2000; Saino et al. 2002; Ludwigs and Becker 2007; Hatch and Westneat 2008; Ismar et al. 2010). Again there are only a few exceptions where the pair bond length is the focus of the study (Ollason and Dunnet 1978; Bradley et al. 1995; Briëd and Jouventin 1999; Pyle et al. 2001; van de Pol et al. 2006; Balbontín et al. 2007; Naves et al. 2007; Nisbet and Dann 2009).

In this study we use a dataset collected from a long-lived seabird, the common tern (Sterna hirundo), to examine the partner choice mechanism and the effect of pair bond duration on breeding performance. Mate retention in the common tern is high but divorces still occur (González-Solís et al. 1999a, c; Becker et al. 2001). Previous studies on this species do not give a concise picture of the effect of partner change. The consequences for first time breeders of changing partner in the subsequent season in terms of breeding success has been studied by Ludwigs and Becker (2007) who found an improvement in the number of fledglings after divorce and a higher breeding success for newly formed pairs after divorce compared to faithful pairings. This suggests a fitness benefit to switching partners although they argue that it is only a result of the earlier arrival of the male in asynchronously arriving pairs. In contrast, González-Solís et al. (1999a) reported no advantage to changing mates: the breeding performance did not change from before to after the divorce and is not different for divorced and faithful pairs. Ludwigs and Becker (2007) gave some evidence that there might also not be an advantage of re-mating with a previous partner. They compared the breeding success of the very selected group of faithful second-time breeders to their first breeding attempt but did not look at more experienced birds. There are no studies for the common tern that study the pair bond length as a continuous increasing function rather than comparing new and old pairings or just first and second-time breeders. There are also no studies on the effect of pair bond length on partner choice for the common tern.

Fowler (1995) highlights the importance of accounting for pair bond length and age in an analysis as single factor analyses can potentially be confounded by the missing factor. Therefore, we would like to know if there is, in addition to a potential influence of pair bond length to reproductive performance, an effect of getting physiologically older. Reproductive performance as a function of age in birds
and mammals tends to be hump shaped with an increase at early reproductive ages, reaching a maximum at intermediate ages and often declining at older ages (Chutton-Brock 1988; Newton 1989; Forslund and Pärt 1995; Rattiste 2004; Nussey et al. 2006, 2009; Weladji et al. 2006, 2010; Balbontín et al. 2007; McCleery et al. 2008; Bouwhuis et al. 2009; Brown and Roth 2009). This age pattern of reproductive performance has also been shown previously in the common tern (Ezard et al. 2007; Rebke et al. 2010).

Due to the injection of automatically read transponders the recapture probability of the marked common tern individuals is close to one in this study (Szostek and Becker 2011). This makes it possible to follow individuals over the course of the study with minimal recorded error. Further, the long-term characteristic of the study ensures that individuals of many age classes and over many different years are included. Despite the high quality of the data, there is missing information that we need to deal with in the analysis. Unknowns in studies from free-living populations include e.g. parts of the resighting history, unknown partners, unknown dates of birth and death, unknown age, unknown sex, missing information before the start of the study for individuals that were already alive. In our particular study it is not always obvious if unmarked partners are new or not. A few individuals were marked as adults and therefore part of their breeding histories are not recorded. For these birds it is unclear if a current partner has already paired with the observed marked bird before the study started. Further, for unmarked partners and also for some of the marked individuals, age, sex and the exact date of death is unknown. Although we know when a marked bird is present in the colony, due to the high recapture probability within the colony, individuals are only rarely found dead. There might also be some individual heterogeneity in reproductive performance, which should be included into the modelling procedure.

Apart from the questions of the effects of age and pair bond length on the breeding output and the mechanism of partner choice we will address several other issues spanning over different hierarchical levels. What does the survival function look like? What is the probability that an adult bird is present in the breeding colony? Do males and females show different patterns in reproductive performance if they are modelled simultaneously? Are there important differences between individuals? What is the probability of breeding repeatedly within a breeding season? Does it depend on previous breeding success?
3.3. Methods

Handling unknowns, and modelling the complicated processes driving reproduction, requires a complex model. A graphical representation of the structure of our model can be found in figure 3.1. We use a hierarchical Bayesian model of the steps that lead to the number of fledglings: a) Is the bird alive or dead? b) If the bird is alive, is it present or absent from the breeding colony? If the bird is present, does it breed? And if the bird breeds, who is the partner? Each bird has a multinomial choice of previous and new available partners. The choice of both partners has to match in order to form a pair. The function for partner choice depends on the cumulative number of past fledglings with a partner and the time spent with that partner. c) What number of fledglings does the pair have in any given breeding occasion within a breeding season? The breeding output depends on the ages of the male and the female, individual effects for the male and the female and the pair bond length. This functional form for the number of fledglings is the same for all three possible breeding occasions within a breeding season. d) Does the pair decide to breed a second time or even a third time in that breeding season? The probability is modelled as dependent on breeding success on the previous occasion within the same season. The first processes in the model happen at the individual level (steps a and b), while the later ones occur at the pair level (steps c and d).

The conditioning of the steps in our model on each other and the modelling of the processes at the individual and the pair level lead to the hierarchical structure. We used a Bayesian approach because estimating the parameters in a maximum likelihood framework for such a high dimensional likelihood surface is very difficult. Using a Bayesian framework allows you to split the model into simpler parts. It is further possible to estimate and impute missing values in this framework directly.

3.3.1. Study Details

We use life history information for common terns collected between 1992 to 2009, defined respectively as $t_1$ and $t_T$. The duration of the study is consequently $T = 18$. The dataset includes $N = 1340$ individually identified birds. The data were collected in a monospecific colony of common terns nesting on six artificial islands in the Banter See in Wilhelmshaven (German North Sea coast, $53^\circ 27'N$, $08^\circ 07'E$). The islands are $10.7 \times 4.6$ m and are surrounded by a 60-cm high wall. The islands are arranged in a row and are $0.9$ m apart.

The terns reproduce in a discrete 3-month long interval every year between mid-May and mid-August. Every 2-3 days during the breeding season the colony in the
Figure 3.1.: The structure of the model. It shows the hierarchical steps that lead to the number of fledglings, where the letters label the steps referred to in the text (a – survival analysis and presence in the colony, b – partner choice, c – number of fledglings per breeding attempt, d – additional breeding attempts). The parts of the model without a coloured background are processes at the single individual level. The yellow background highlights the part of the model that depends on the decision of both partners. The blue background represents parts that are modelled at the pair level.

Banter See has been checked for new eggs, hatched and fledged chicks. The details of this ongoing long-term study are described in Becker et al. (2001).

Since 1992, all chicks in the colony in the Banter See have been ringed and marked with subcutaneously implanted transponders (TROVAN ID 100; TROVAN, Köln, Germany) after hatching when they are about 14-18 days old (Becker and Wendeln 1997). For these birds the exact age is known. A chick that reached an age of 18 days that was not found dead during the remainder of the breeding season was defined as fledged. The sex of all chicks marked with a transponder has been determined by standard molecular methods since 1998 (Becker and Wink 2002) and before by behavioural observations. If a bird of unknown sex paired with a bird of known sex
at some point during the study it was automatically assigned the opposite sex. In addition to the chicks some breeding adults were marked with transponders between 1992 and 1995. As common terns have been ringed since 1980 at other places in Wilhelmshaven and when the colony started to establish in 1984, at the Banter See, age information for some of these birds is available. There are 50 birds that were implanted with transponders as breeding adults and for which the year of birth is not known. Thirteen birds that appear in the dataset as a partner of a tern with a transponder are identified by ring number only. These partners were identified by sightings. For 11 of these individuals the year of birth is available but the complete breeding histories for all these birds are not known. These birds identified by ring number are handled in the analysis for their own breeding history like unknown birds but with a known age if the birth date is available and like a known bird for the breeding history of their marked partners. As only chicks and 101 breeders have been PIT-tagged since 1992, there are birds at the colony that do not have an implanted transponder; they are typically immigrants.

The individual specific alphanumeric code from the injected transponders is read automatically every 5-10 seconds by antennae on 44 elevated resting platforms for terns within a distance of $\leq11$ cm. This method ensures the identification of all PIT-tagged adults present in the colony. To identify the two parent terns belonging to each nest additional antennae are placed around every incubated clutch temporarily for 1-2 days during the breeding season. This automatic detection system is described in detail in Becker et al. (2001, 2008b). It ensures a recapture probability very close to one during the breeding season, when the bird is present in the Banter See colony (Becker et al. 2008b; Szostek and Becker 2011). There is no information on appearance in other, neighbouring, colonies. Since birds are not always present in the colony we include a presence history into the analysis equivalently to the typically used recapture history in studies of free-living populations. A bird is automatically classified as alive, when it is present in the colony but it is not automatically classified as dead if not. For the estimation of date at death the presence histories are taken into account.

In our model we incorporate information on the bird being alive or dead, being present in the colony, the decision to breed up to three times within a breeding season, the partner choice and ultimately the number of fledglings. Although we have a number of sub-functions for the different processes we combine them into one model and estimate them simultaneously in a hierarchical Bayesian framework, rather than by conducting a separate analysis for each process.
3.3.2. Assignment of the Partner Status

Challenges for our study are that not all of the partners are marked and not all of the birds have their entire breeding histories recorded. Therefore it is not always obvious if unmarked partners are new or not. An important step in data preparation is the assignment of partner status for all individuals with a transponder. The assignment of partner status is conducted using rules displayed in the flow chart (figure 3.2). If a bird is alive, present in the colony and breeding it needs to have a partner. The partner is either new (0), a previous mate (1) or the status of the partner is not known (2). The first breeding time can be known or unknown. In both cases the current partner is marked or unmarked. If the first breeding time is recorded the partner is always new if the current breeding event is the first. In all other cases the current partner needs to be compared to the list of previous partners and during the estimation procedure to the list of potential previous partners to assign its status.

When a bird breeds once, or repeatedly, with an unmarked bird and with a marked bird in the following year that was only marked as an adult in that following year, we assume that the partner is the same bird. It appears as unmarked before it got the transponder as an adult and as marked afterwards. There are only 7 pairs where that occurs.

For the 50 birds where we do not have the entire breeding history we assume for simplicity that when an unmarked or a marked partner appears for the first time in the observed history it is a new partner. Although, in theory, that partner could have already appeared in the breeding history before we started observing.

3.3.3. Model Details

The events we are modelling here are inherently hierarchical; from survival to producing fledglings, each process is conditioned on another process at a higher hierarchy. For instance, the choice of an individual to breed in a given year is conditioned on surviving to that year; if the bird chooses to breed then it has to select a partner, while its choice of partner can potentially influence the breeding output of the pair (figure 3.1). Moreover, these processes occur at different levels; from survival to choosing to breed, the decisions are made by each individual (parts without a coloured background in figure 3.1). The choice of partner also happens at the individual level, but the choice of two birds has to match (the part highlighted in yellow in figure 3.1). Producing fledglings is the result of pair level processes (i.e. involving both parents, the part of the model highlighted in blue in figure 3.1). This bilevel-hierarchical structure, along with the amount of missing information, can be
Figure 3.2.: Flowchart of the assignment of partner status: 0 – new partner, 1 – previous partner, 2 – unknown if new or previous partner. In order to assign the partner status the following questions are asked: Is the 1st breeding time and thus the entire breeding history known? Is the current partner marked or unmarked? Is the current breeding event the bird’s first time breeding? Then the current partner ID is compared to the list of all previous partners. At the end certain assumptions are made to simplify the model (the parts in red).

analysed under a Bayesian framework. Our Bayesian model combines recent developments in survival analysis (Colchero and Clark 2011) with a mixture of generalised linear models, each of which corresponds to a different hierarchy in figure 3.1.

According to the graphical depiction of the processes in figure 3.1, we divide the model into four main sections, all of which are conditioned on the preceding section: a) age-specific survival analysis, and presence in the colony; b) choosing a partner; c) producing fledglings; d) additional breeding attempts. Below we explain each section in detail.
a) Survival Analysis and Presence in the Colony

The first level in our model is the analysis of survival patterns from one age to the next. The data consist of individual capture-recapture histories, reflecting the presence history, bound by times of birth for most tagged individuals, and times of death for a few of them. During the modelling procedure we estimate the population-level survival parameters, and also the latent (unknown) times of birth and death when missing. The times of birth and death need to be estimated for each tagged individual for which this information is missing and each partner without a tag and consequently no information on birth and death. We use a modelling framework for capture-recapture data recently developed by Colchero and Clark (2011) which combines models for survival and recapture probability accounting for right-censoring and left-truncation, that imputes latent states (i.e. unknown times of birth and death). Colchero and Clark (2011) show that the model provides accurate estimates of survival patterns and unknown times of birth and death by applying it to simulated data and to a capture-recapture dataset of Soay sheep (Ovis aries).

Let $x$ represent age (in years) and $X$ be the random variable for age at death. Our modelling approach requires the definition of a parametric model for mortality rate which, scaled to a time interval of length $dx$, can be represented as $\mu(x|\theta) = \Pr\{x < X < x + dx|X > x; \theta\}/dx$, where $\theta$ are parameters to be estimated. We assume that individuals on average die in the middle of an age interval (i.e. $dx/2$). We use a Gompertz hazard (Gompertz 1825) to describe the underlying mortality over age $x$:

$$\mu(x|\alpha, \beta) = \alpha e^{\beta x}, \quad (3.1)$$

where $\alpha$ is the baseline mortality and $\beta$ is the rate parameter. We use this Gompertz hazard function to calculate the survival function:

$$S(x|\alpha, \beta) = e^{\frac{\alpha}{\beta}(1-e^{\beta x})}, \quad (3.2)$$

and the probability density function of ages at death:

$$f(x|\alpha, \beta) = S(x)\mu(x) = \alpha e^{\beta x + \frac{\alpha}{\beta}(1-e^{\beta x})}, \quad (3.3)$$

with $\alpha > 0$ and $-\infty < \beta < \infty$. We ran a separate preliminary maximum likelihood survival analyses with a Gompertz, Gompertz-Makeham, Weibull, and Siler function using only the known-age individuals. Although the Gompertz-Makeham showed the best performance in this preliminary test for simplicity we chose the Gompertz...
function as the performance of all models was similar with a difference in the Akaike Information Criterion (AIC) between the best and worst performing models of 4.9 (see 3.5.3 for discussion).

Although recapture probabilities are essentially equal to 1 within the breeding colony (see section 3.3.1), detection is still conditioned on the bird being present in the colony. Although there is no information on attendance at other breeding colonies, the common tern is known to show high site fidelity after the first breeding event (del Hoyo et al. 1996) and therefore the majority of individuals not present at the colony in intermediate years are most likely skipping a breeding season. Thus, analogous to a typical capture-recapture dataset, we note the presence at the colony of an individual $i$ at time $t$ with the indicator $c_{i,t}$, which assigns 1 if the individual is present and 0 otherwise. The probability of being present in the colony is a binomial process represented by $c_{i,t}$, which depends on the individual still being alive at that age:

$$p(c_{i,t}|x_{i,t} < X_i) = \text{Bin}(\pi),$$

(3.4)

where the parameter $\pi$ is the probability of being present at the colony. We represent an individual $i$’s ‘presence history’ (analogous to the capture history) with the vector $c_i$, while $C$ is the matrix for all histories in the dataset. We use a conjugate Beta distribution to sample $\pi$ directly:

$$p(\pi) = \text{Beta}(a_1 + \sum(o_i), a_2 + \sum(T_i - o_i)),$$

(3.5)

where $a_1$ and $a_2$ are the hyperparameters for the Beta conjugate prior for $\pi$, $o_i$ is the number of years bird $i$ was observed in the colony and $T_i$ is the total number of years individual $i$ could have been detected, which we calculate as

$$T_i = \min(d_i, t_T) - \max(b_i, t_1),$$

where $b_i$ and $d_i$ are years of birth and death for individual $i$ if observed.

Under our Bayesian framework, we can estimate the posterior of all unknowns, $p(\alpha, \beta, \pi, X_0|X_1, C)$, where $X_0$ and $X_1$ are the subsets of unknown and of known ages at death, respectively, and $X_0 \cup X_1 = X$. The model estimates the conditional posterior for the parameters, conditioned on the data and priors, and the conditional posterior of the latent states (i.e. $X_0$) conditioned on the mortality model, the probability of the bird not being detected for the number of years between the last detection and the time of death (data model) and the prior for the age distribution.
Since we assume that no bird comes back to the breeding colony before age two, we conditioned our age-specific survival analysis on individuals that were at least that age. Common terns usually start visiting the breeding colonies after spending their first two years in Africa. At that age they are commonly prospectors and many of them become breeders at a later age (Ludwigs and Becker 2002; Dittmann and Becker 2003; Becker et al. 2008a). There were only two birds in the entire study that were already present in the colony at age one. Birds of unknown age are assumed to be at least three years old when we observe them breeding for the first time as most common terns start breeding at an age of three years (Becker et al. 2001, 2008a; Ludwigs and Becker 2002).

**b) Partner Choice**

A bird $i$ that has survived to year $t$ and is present in the colony at that time (i.e. $c_{i,t} = 1$) has the choice to breed. Although a few young birds did not breed despite being present at the colony, most birds did attempt to breed. We do not model the probability of breeding explicitly, but all following steps are conditioned on breeding (i.e. $y_{i,t,1} = 1$) in the first breeding occasion $k = 1$ within a season.

At this point, each breeding bird has to find a mate. Although we assume that the decision of which partner to choose is made by each individual independently, the choice of both partners has to match. We define $v_{i,t}$ as the choice of partner that bird $i$ makes at time $t$, and $v_{j,t}$ is the choice that bird $j$ makes. Consequently, the joint probability of $v_{i,t} = j$ and $v_{j,t} = i$ is simply the probability that partner $i$ chooses partner $j$ times the probability that partner $j$ chooses partner $i$, which yields:

$$p(v_{i,t} = j, v_{j,t} = i) = p(v_{i,t} = j) \times p(v_{j,t} = i). \quad (3.6)$$

The partner choice, $v_{i,t}$, is a multinomial process that includes past mates that are still alive and all other birds of the opposite sex present in the colony at time $t$ (i.e. potential new partners), for a total of $L_t$ available individuals. The number of new partners is given by the difference between the number of nests in that breeding season and the number of previous partners that are still alive. We assumed that all birds choose partners at the same time and from the same pool of birds. This is of course not entirely consistent with the actual partner choice process in seabirds but, for simplicity we did not include further covariates such as arrival time into the model. Still, this multinomial choice is informed by the bird’s past history with
previous partners. Thus, the probability that bird \( i \) chooses partner \( j \) is given by:

\[
p(v_{i,t} = j | c_{i,t} = 1, y_{i,t,1} = 1) = \frac{e^{\phi_0 + \phi_1 Z_{i,j,t} + \phi_2 T_{i,j,t}}}{\sum_{l=1}^{L_t} e^{\phi_0 + \phi_1 Z_{i,l} + \phi_2 T_{i,l,t}}},
\]

(3.7)

where the \( \phi \)s are the link function parameters for the probability of choosing a certain partner, \( Z_{i,j,t} = \sum_{\tau=t_1}^{t-1} z_{i,j,\tau} \) is the total number of fledglings produced by partners \( i \) and \( j \) before year \( t \), and \( T_{i,j,t} \) is the number of times they have been together in the past. Thus, for birds that have not been together on previous occasions, both \( Z_{i,j,t} \) and \( T_{i,j,t} \) will be equal to 0, making \( \phi_0 \) the baseline parameter for new partners. This function is flexible and allows us to consider the process of deciding for an unfamiliar partner separately from deciding for a familiar mate based on past experience.

Here arises a complication when marked birds breed with unmarked ones. In that case, the model needs to treat the unmarked birds as potential partners and assign them an ID code. We initialise the model by assuming that all unmarked birds are different birds, each with individual ID codes. For example, within a dataset of \( N = 10 \) birds and study duration of \( T = 4 \) years, the marked bird number \( i = 1 \) breeds only with unmarked birds. The vector that corresponds to that bird’s partner history should be \( v_1 = [\text{NA}, \text{NA}, \text{NA}, \text{NA}] \), where NA refers to the lack of an ID code (i.e. transponder or band) of the partner. Since we are assigning ID numbers to these unknown partners, the partner history is then initialised as \( v_1 = [11, 12, 13, 14] \) (i.e. starting from \( N + 1 \)). However, based on the relationships described above, our algorithm might impute that the partner at times \( t = 2 \) and \( t = 3 \) should be bird 11 again, which yields \( v_1 = [11, 11, 11, 14] \). We assume that all unmarked bird ID codes that are initially assigned to a specific marked bird will only breed with that bird or will be eliminated from the dataset. Since the breeding output between partners depends on the age of both partners and only living birds can pair again, times of birth and death for the unmarked bird have to be assigned. During the modelling procedure it is important to keep track of the age, the number of times this bird breeds with its marked partner and the breeding success. The inclusion of covariates in the multinomial decision informs the assignment of whether an unmarked partner is new or not.

c) Number of Fledglings per Breeding Attempt

After forming the pairs, the model moves to the nest level (parts highlighted with a blue background colour in figure 3.1). Both partners either marked or generated from the model are matched and assigned to a nest \( m \) at time \( t \). We denote the total number of nests as \( M \). At this point we will use the subindex \( m \) as equivalent to the
index $i, j$ used in the previous section, which means that each nest $m$ corresponds to a specific combination of parents $i$ and $j$, and each time the same parents breed together their nest has the same $m$ assigned. By switching from the individual to the nest level we avoid the problem of repeating the number of fledglings for both partners within a pair.

Within one breeding season, common terns can attempt to lay eggs on several consecutive occasions $k$ with a maximum of $K = 3$ occasions. At each occasion $k$, the number of fledglings produced by nest $m$, $z_{m,t,k}$, can range from zero to three. For simplification we assume that the processes that determine $z_{m,t,k}$ are the same for all attempts during one breeding season $t$. More than three fledglings per breeding occasion never appeared in the dataset and therefore this upper bound needs to be incorporated into the model. We use a Poisson distribution truncated at an upper value of three to model the number of fledglings for nest $m$ at time $t$ and occasion $k$:

$$ p(z_{m,t,k}) = \text{Pois}^3(z_{m,t,k}|\lambda_{m,t,k}), \quad (3.8) $$

where $\lambda_{m,t,k}$ is the expected number of fledglings for nest $m$ at time $t$ and occasion $k$. The truncated Poisson distribution is calculated as

$$ \text{Pois}^3(z_{m,t,k}|\lambda_{m,t,k}) = \frac{\text{Pois}(z_{m,t,k}|\lambda_{m,t,k})}{\sum_{l=0}^{3} \text{Pois}(l|\lambda_{m,t,k})}, \quad \lambda_{m,t,k} \in [0, 3], \quad (3.9) $$

where the denominator in equation 3.9 is the Poisson cumulative density function (CDF), $P\{l \leq 3|\lambda_{m,t,k}\}$. In our analysis the number of fledglings per nest is not overdispersed; the mean and the variance are very close with 0.66 and 0.68 respectively. Thus, the truncated Poisson distribution is a reasonable assumption. The link function for the expected number of fledglings is given by:

$$ \ln(\lambda_{m,t,k}) = \omega_0 + \omega_1 T_{m,t} + \sum_{s=1}^{2} [\omega_{2,s}(x_{m,s,t} + \omega_{3,s})^2 + u_{m,s}], \quad (3.10) $$

with $-\infty < \omega_0 < \infty$, $-\infty < \omega_1 < \infty$, $\omega_{2,s} \leq 0$ and $-\infty < \omega_{3,s} < \infty$. The parameter $\omega_0$ determines the upper bound of $\ln(\lambda_{m,t})$, while parameter $\omega_1$ relates the number of fledglings with the number of years both parents had nested together before year $t$, represented by $T_{m,t}$ (equivalent to $T_{i,j,t}$ in equation 3.7). The subindex $s$ refers to the sex of the bird, where $s = 1$ if the bird is a male and $s = 2$ otherwise. The quadratic term in equation 3.10 allows us to test non-linear effects on the number of fledglings as a function of the parents’ ages for each sex; the parameters $\omega_{2,s}$
Table 3.1.: Variables, symbols and their definition

<table>
<thead>
<tr>
<th>Variable / Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>Individual</td>
</tr>
<tr>
<td>j</td>
<td>Other individual</td>
</tr>
<tr>
<td>s</td>
<td>Sex, with males being defined as $s = 1$ and females as $s = 2$</td>
</tr>
<tr>
<td>N</td>
<td>Total number of marked birds</td>
</tr>
<tr>
<td>L</td>
<td>Total number of available partners</td>
</tr>
<tr>
<td>m</td>
<td>Nest</td>
</tr>
<tr>
<td>M</td>
<td>Total number of nests</td>
</tr>
<tr>
<td>t</td>
<td>Year</td>
</tr>
<tr>
<td>T</td>
<td>Duration of the study</td>
</tr>
<tr>
<td>$T_i$</td>
<td>Total number of years individual $i$ could have been detected</td>
</tr>
<tr>
<td>$a_i$</td>
<td>Total number of years bird $i$ was observed in the colony</td>
</tr>
<tr>
<td>$b_i$</td>
<td>Year of birth for individual $i$ if observed</td>
</tr>
<tr>
<td>$d_i$</td>
<td>Year of death for individual $i$ if observed</td>
</tr>
<tr>
<td>$k$</td>
<td>Occasion</td>
</tr>
<tr>
<td>$K$</td>
<td>Maximum number of occasions</td>
</tr>
<tr>
<td>$x$</td>
<td>Age</td>
</tr>
<tr>
<td>$dx$</td>
<td>Length of age interval</td>
</tr>
<tr>
<td>$X$</td>
<td>Vector with ages at death for all individuals</td>
</tr>
<tr>
<td>$X_0$</td>
<td>Vector with unknown ages at death</td>
</tr>
<tr>
<td>$X_1$</td>
<td>Vector with known ages at death</td>
</tr>
<tr>
<td>$\mu()$</td>
<td>Mortality hazard</td>
</tr>
<tr>
<td>$S()$</td>
<td>Survival function</td>
</tr>
<tr>
<td>$f()$</td>
<td>Probability density function</td>
</tr>
<tr>
<td>$c$</td>
<td>Indicator for presence in the colony</td>
</tr>
<tr>
<td>$c_i$</td>
<td>Vector of the entire presence history for individual $i$</td>
</tr>
<tr>
<td>$C$</td>
<td>Matrix of all presence histories</td>
</tr>
<tr>
<td>$a_1, a_2$</td>
<td>Hyperparameters for the Beta conjugate prior for the probability of being present in the colony</td>
</tr>
<tr>
<td>$v$</td>
<td>Choice of partner</td>
</tr>
<tr>
<td>$v_i$</td>
<td>Vector of all partner choices for individual $i$</td>
</tr>
<tr>
<td>$V_0$</td>
<td>Vector with all choices for unknown partners</td>
</tr>
<tr>
<td>$V_1$</td>
<td>Vector with all choices for known partners</td>
</tr>
<tr>
<td>$z$</td>
<td>Number of fledglings produced</td>
</tr>
<tr>
<td>$Z$</td>
<td>Matrix of all histories of number of fledglings</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>Expected number of fledglings</td>
</tr>
<tr>
<td>$g$</td>
<td>Indicator for breeding</td>
</tr>
<tr>
<td>$Y_k$</td>
<td>Matrix of all breeding histories for occasion $k$</td>
</tr>
<tr>
<td>$r$</td>
<td>Indicator of breeding success on the previous occasion</td>
</tr>
<tr>
<td>$\theta$</td>
<td>Probability of breeding an additional time within the same year</td>
</tr>
<tr>
<td>$Z_{i,j,t}$</td>
<td>Total number of fledglings produced by partners $i$ and $j$ before year $t$</td>
</tr>
<tr>
<td>$T_{i,j,t} = T_{m,t}$</td>
<td>Number of times bird $i$ and $j$ had nested together in the past</td>
</tr>
</tbody>
</table>

control the steepness of the change in breeding performance with age, while the parameters $\omega_{3,s}$ control the age at which the highest performance is reached. If $\omega_{3,s} < 0$, the breeding performance declines with age, if $\omega_{3,s} \in [0, x_{\text{max}})$, where $x_{\text{max}}$ is the maximum lifespan of the species, then breeding performance is reached at the age indicated by $\omega_{3,s}$, while if $\omega_{3,s} > x_{\text{max}}$ the breeding performance increases constantly with age.

The parameters $u_{m,s}$ are individual effects that facilitate testing for heterogeneity in breeding performance. One fixed individual specific effect is estimated for each bird. There is no specific underlying distribution assumed for these individual intercepts. The range of values for these parameters is $u_{m,s} \in [-1, 0]$; good breeders will have a $u_{m,s}$ close to 0, while for poor breeders the parameter should tend to
−1. If both parents are extremely poor breeders (both parents have an individual effect of -1), the expected number of fledglings will be only 15% that of a nest with two extremely good breeders (both parents have an individual effect of 0) when the parameter estimates of the model are used for the prediction. This is very similar to the maximum difference in reproduction of 15.2% caused by quality in a study of the little penguin (Eudyptes minor, Nisbet and Dann 2009). Figure 3.6 shows the shift of the intercept in the link function due to the individual effect (left hand panel) and the translation into the corresponding functions of the number of fledglings over age (right hand panel). In addition, by including these individual effects we control for autocorrelation between repeated readings from the same individuals.

We assume that partners remain together for the entire duration of the breeding season and thus there is no change in pairings between occasions (González-Solís et al. 1999c; Ludwig and Becker 2008b). Although individuals could potentially change their partner within a breeding season, we only recorded a negligible number of five cases.

**d) Additional Breeding Attempts**

The decision to breed at additional occasions in one breeding season is made at the nest level. We denote as $y_{m,t,k}$ the event in which parents from nest $m$ at time $t$ attempt to breed on a second and third occasion (i.e. $k \in \{2,3\}$). The decision to breed an additional time is a binomial process:

$$p(y_{m,t,k}|c_{m,t} = 1, y_{m,t,k-1} = 1) = \text{Bin}(y_{m,t,k} | \theta_{m,t,k}) \quad k \in \{2,3\}. \quad (3.11)$$

We model the probability of breeding a second and a third time in one year as being dependent on whether the pair bred successfully on the previous occasion within the same year $r_{m,t,k-1}$, $r_{m,t,k-1} = 0$ if $z_{m,t,k-1} = 0$ and $r_{m,t,k-1} = 1$ if $z_{m,t,k-1} > 0$. The link function is given by:

$$\text{logistic}(\theta_{m,t,k}) = \gamma_{0,k} + \gamma_{1,k} r_{m,t,k-1} \quad k \in \{2,3\}. \quad (3.12)$$

A summary of all variables and symbols in the model is given in table 3.1 and table 3.2 lists all parameters estimated.
3.3.4. Bayesian Modelling

The joint posterior for all unknowns in our model is:

\[
p(\alpha, \beta, \pi, X_0, \Phi, V_0, \Omega, \Gamma | X_1, C, V_1, Z_k, Y_k) \propto p(X | \alpha, \beta) p(C | \pi, X) \times p(V | \Phi) \times \prod_{k=1}^{3} p(Z_k | \Omega) \times \prod_{k=2}^{3} p(Y_k | \Gamma) \times p(\alpha, \beta) p(\pi) p(\Phi) p(\Omega) p(\Gamma),
\]

where matrix \( V \) of dimension \( N \times T \) represents all the partner choices in the dataset and \( V_0 \) and \( V_1 \) are the subsets of unknown and known partners, respectively. Variables \( Z_k \) and \( Y_k \) are \( M \times T \) matrices for the number of fledglings and the indicator for breeding on occasion \( k \), respectively. The elements \( \Phi, \Omega \) and \( \Gamma \) are the vectors for parameters, \( \phi, \omega \) and \( \gamma \), in equations 3.7, 3.10 and 3.12, respectively. The labels above the braces in equation 3.13 show the section of the model that each element corresponds to.

We use uninformative priors except for survival, where we use a Gompertz function. Colchero and Clark (2011) show that the prior choice for the survival sub-model does not influence the parameter estimates significantly.

Our hierarchical framework needs only the conditionals for the posterior simulated by Gibbs sampling, specifically Metropolis-within-Gibbs (Gelfand and Smith 1990; Clark 2007). This Monte Carlo Markov Chain (MCMC) procedure consists of proposing parameters and latent states at each step, and accepting or rejecting them with some probability. Thus, in each iteration the missing value is conditioned on the previous parameter estimates and the data.

To specify how many steps are needed before convergence (i.e. burn-in sequence), we ran 10 parallel MCMC sequences with different randomly generated starting values. From the resulting MCMC chains, we computed potential scale reduction factors for each parameter (Gelman et al. 2004). This diagnostic is calculated as 
\[
\hat{R} = \sqrt{\hat{v}^+/W},
\]
where \( W \) is a measure of the within-sequence variance and \( \hat{v}^+ \) is a weighted average of the between-sequence variance (\( B \)) and \( W \). Convergence is attained when \( \hat{R} \) is close to 1. The resulting potential scale reduction shows that
Table 3.2.: Estimated parameters and their definition

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>Gompertz baseline mortality</td>
</tr>
<tr>
<td>$\beta$</td>
<td>Gompertz rate parameter</td>
</tr>
<tr>
<td>$\pi$</td>
<td>Probability of being present in the colony</td>
</tr>
<tr>
<td>$\phi_0$</td>
<td>Baseline link function parameter of new partners for the probability of choosing a certain partner</td>
</tr>
<tr>
<td>$\phi_1$</td>
<td>Link function parameter that relates the accumulated number of fledglings produced in the past to the probability of choosing a certain partner</td>
</tr>
<tr>
<td>$\phi_2$</td>
<td>Link function parameter that relates the accumulated number of times breeding together in the past to the probability of choosing a certain partner</td>
</tr>
<tr>
<td>$\Phi$</td>
<td>Vector of all link function parameters for the probability of choosing a certain partner</td>
</tr>
<tr>
<td>$\omega_0$</td>
<td>Baseline link function parameter for the number of fledglings, which determines the upper bound of the logarithmic expected number of fledglings</td>
</tr>
<tr>
<td>$\omega_1$</td>
<td>Link function parameter that relates the number of years both parents had nested together in the past to the number of fledglings</td>
</tr>
<tr>
<td>$\omega_{2,s}$</td>
<td>Link function parameter for males and females to control the steepness of the change in number of fledglings with age</td>
</tr>
<tr>
<td>$\omega_{3,s}$</td>
<td>Link function parameter for males and females to control the age at which the highest number of fledglings is reached</td>
</tr>
<tr>
<td>$\Omega$</td>
<td>Vector of all link function parameters for the number of fledglings</td>
</tr>
<tr>
<td>$u$</td>
<td>Individual effects for the number of fledglings</td>
</tr>
<tr>
<td>$\gamma_{0,k}$</td>
<td>Link function parameter for the probability of breeding on occasion $k$ after a previous unsuccessful occasion $k - 1$</td>
</tr>
<tr>
<td>$\gamma_{1,k}$</td>
<td>Link function parameter for the additional effect of a previous successful breeding occasion $k - 1$ on the probability of breeding on occasion $k$</td>
</tr>
<tr>
<td>$\Gamma$</td>
<td>Vector of all link function parameters for the probability of breeding an additional time</td>
</tr>
</tbody>
</table>

Convergence is attained after the first 20,000 steps and that serial autocorrelation is negligible with a lag of 100 iterations. We ran all MCMC algorithms for a total of 50,000 iterations. The predicted parameter values are given with the 95% credible interval (CRI).

We use the deviance information criterion (DIC) to compare the model with the individual effects to the model without these effects (Spiegelhalter et al. 2002). We wrote the model with the statistical software R version 2.12.0 (R Development Core Team 2010).

### 3.4. Results

All parameters in the model with the individual effects and the model without the individual effects converged (see $\hat{R}$ in table 3.3 and table 3.4). Parameters not directly connected to the number of fledglings $\alpha$, $\beta$, $\pi$ and $\Gamma$ are similar in both models, while many parameters related to the number of fledglings $\Phi$ and $\Omega$ show differences (see table 3.3 and table 3.4 for parameter values). The model with the individual effects on reproductive output fits the data much better than the model without the individual effects, with a DIC of 59371 compared to a DIC of 59841, respectively. Thus, we only report the results of the model with the individual effects in the following unless stated otherwise.
Table 3.3.: Parameter values for the model with individual effects

<table>
<thead>
<tr>
<th>Parameter</th>
<th>50%</th>
<th>2.5%</th>
<th>97.5%</th>
<th>B</th>
<th>W</th>
<th>( \hat{v}^+ )</th>
<th>( \hat{R} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )</td>
<td>0.1109</td>
<td>0.1002</td>
<td>0.1221</td>
<td>0.0020</td>
<td>0.0000</td>
<td>0.0000</td>
<td>1.0011</td>
</tr>
<tr>
<td>( \beta )</td>
<td>0.0832</td>
<td>0.0691</td>
<td>0.0975</td>
<td>0.0044</td>
<td>0.0001</td>
<td>0.0001</td>
<td>1.0014</td>
</tr>
<tr>
<td>( \pi )</td>
<td>0.8925</td>
<td>0.8847</td>
<td>0.9002</td>
<td>0.0009</td>
<td>0.0000</td>
<td>0.0000</td>
<td>1.0009</td>
</tr>
<tr>
<td>( \gamma_{0,2} )</td>
<td>-1.3600</td>
<td>-1.4860</td>
<td>-1.2389</td>
<td>0.5485</td>
<td>0.0040</td>
<td>0.0040</td>
<td>1.0023</td>
</tr>
<tr>
<td>( \gamma_{1,2} )</td>
<td>-2.8619</td>
<td>-3.3434</td>
<td>-2.4362</td>
<td>6.1579</td>
<td>0.0516</td>
<td>0.0516</td>
<td>1.0020</td>
</tr>
<tr>
<td>( \gamma_{0,3} )</td>
<td>-3.6236</td>
<td>-4.4142</td>
<td>-2.9185</td>
<td>10.4550</td>
<td>0.1395</td>
<td>0.1395</td>
<td>1.0012</td>
</tr>
<tr>
<td>( \gamma_{1,3} )</td>
<td>-1.8100</td>
<td>-2.6565</td>
<td>-0.1551</td>
<td>41.427</td>
<td>0.5209</td>
<td>0.5216</td>
<td>1.0013</td>
</tr>
<tr>
<td>( \phi_{0} )</td>
<td>1.0581</td>
<td>-60.4292</td>
<td>59.5134</td>
<td>280821.0123</td>
<td>938.2273</td>
<td>947.5564</td>
<td>1.0050</td>
</tr>
<tr>
<td>( \phi_{1} )</td>
<td>-0.1975</td>
<td>-2.1939</td>
<td>2.0754</td>
<td>96.3533</td>
<td>1.2469</td>
<td>1.2500</td>
<td>1.0013</td>
</tr>
<tr>
<td>( \phi_{2} )</td>
<td>2.5772</td>
<td>1.0394</td>
<td>4.5222</td>
<td>110.9331</td>
<td>0.5209</td>
<td>0.5216</td>
<td>1.0013</td>
</tr>
<tr>
<td>( \omega_{0} )</td>
<td>0.7782</td>
<td>0.6203</td>
<td>0.9604</td>
<td>9.3884</td>
<td>0.0073</td>
<td>0.0076</td>
<td>1.0212</td>
</tr>
<tr>
<td>( \omega_{1} )</td>
<td>0.0016</td>
<td>-0.0363</td>
<td>0.0323</td>
<td>0.3052</td>
<td>0.0002</td>
<td>0.0002</td>
<td>1.0207</td>
</tr>
<tr>
<td>( \omega_{2,1} )</td>
<td>-0.0052</td>
<td>-0.0089</td>
<td>-0.0013</td>
<td>0.0016</td>
<td>0.0000</td>
<td>0.0000</td>
<td>1.0074</td>
</tr>
<tr>
<td>( \omega_{2,2} )</td>
<td>11.2863</td>
<td>8.8639</td>
<td>17.7662</td>
<td>6219.1326</td>
<td>4.5456</td>
<td>4.7528</td>
<td>1.0225</td>
</tr>
<tr>
<td>( \omega_{3,1} )</td>
<td>10.6958</td>
<td>8.1607</td>
<td>15.1447</td>
<td>2540.9383</td>
<td>3.0307</td>
<td>3.1153</td>
<td>1.0139</td>
</tr>
</tbody>
</table>

Table 3.4.: Parameter values for the model without individual effects

<table>
<thead>
<tr>
<th>Parameter</th>
<th>50%</th>
<th>2.5%</th>
<th>97.5%</th>
<th>B</th>
<th>W</th>
<th>( \hat{v}^+ )</th>
<th>( \hat{R} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )</td>
<td>0.1109</td>
<td>0.0999</td>
<td>0.1221</td>
<td>0.0027</td>
<td>0.0000</td>
<td>0.0000</td>
<td>1.0014</td>
</tr>
<tr>
<td>( \beta )</td>
<td>0.0830</td>
<td>0.0690</td>
<td>0.0973</td>
<td>0.0041</td>
<td>0.0001</td>
<td>0.0001</td>
<td>1.0013</td>
</tr>
<tr>
<td>( \pi )</td>
<td>0.8925</td>
<td>0.8847</td>
<td>0.9002</td>
<td>0.0009</td>
<td>0.0000</td>
<td>0.0000</td>
<td>1.0009</td>
</tr>
<tr>
<td>( \gamma_{0,2} )</td>
<td>-1.3605</td>
<td>-1.4862</td>
<td>-1.2426</td>
<td>0.4331</td>
<td>0.0039</td>
<td>0.0039</td>
<td>1.0019</td>
</tr>
<tr>
<td>( \gamma_{1,2} )</td>
<td>-2.8588</td>
<td>-3.5305</td>
<td>-2.4396</td>
<td>2.1427</td>
<td>0.5209</td>
<td>0.5216</td>
<td>1.0007</td>
</tr>
<tr>
<td>( \gamma_{0,3} )</td>
<td>-3.6083</td>
<td>-4.4073</td>
<td>-2.9143</td>
<td>8.2402</td>
<td>0.1424</td>
<td>0.1427</td>
<td>1.0009</td>
</tr>
<tr>
<td>( \gamma_{1,3} )</td>
<td>-1.1324</td>
<td>-2.6728</td>
<td>0.1736</td>
<td>24.5837</td>
<td>0.5238</td>
<td>0.5246</td>
<td>1.0008</td>
</tr>
<tr>
<td>( \phi_{0} )</td>
<td>1.5704</td>
<td>-60.3072</td>
<td>61.3100</td>
<td>340141.6139</td>
<td>954.7340</td>
<td>966.0398</td>
<td>1.0059</td>
</tr>
<tr>
<td>( \phi_{1} )</td>
<td>-0.1059</td>
<td>-2.2393</td>
<td>2.1404</td>
<td>174.3834</td>
<td>1.2923</td>
<td>1.2981</td>
<td>1.0022</td>
</tr>
<tr>
<td>( \phi_{2} )</td>
<td>2.5440</td>
<td>0.9889</td>
<td>4.5250</td>
<td>65.9753</td>
<td>0.8752</td>
<td>0.8773</td>
<td>1.0012</td>
</tr>
<tr>
<td>( \omega_{0} )</td>
<td>0.0421</td>
<td>-0.2154</td>
<td>0.2099</td>
<td>14.5860</td>
<td>0.0110</td>
<td>0.0115</td>
<td>1.0219</td>
</tr>
<tr>
<td>( \omega_{1} )</td>
<td>0.0046</td>
<td>-0.0218</td>
<td>0.0321</td>
<td>0.1913</td>
<td>0.0002</td>
<td>0.0002</td>
<td>1.0165</td>
</tr>
<tr>
<td>( \omega_{2,1} )</td>
<td>-0.0041</td>
<td>-0.0081</td>
<td>-0.0009</td>
<td>0.0014</td>
<td>0.0000</td>
<td>0.0000</td>
<td>1.0066</td>
</tr>
<tr>
<td>( \omega_{2,2} )</td>
<td>-0.0040</td>
<td>-0.0079</td>
<td>-0.0010</td>
<td>0.0011</td>
<td>0.0000</td>
<td>0.0000</td>
<td>1.0057</td>
</tr>
<tr>
<td>( \omega_{3,1} )</td>
<td>13.4330</td>
<td>10.1429</td>
<td>25.9333</td>
<td>14720.1169</td>
<td>12.9767</td>
<td>13.4669</td>
<td>1.0187</td>
</tr>
<tr>
<td>( \omega_{3,2} )</td>
<td>12.2911</td>
<td>9.1497</td>
<td>19.0681</td>
<td>5769.4644</td>
<td>6.1870</td>
<td>6.3791</td>
<td>1.0154</td>
</tr>
</tbody>
</table>

The Gompertz mortality and the survival curves with the estimated baseline \( \alpha = 0.1109 \) and rate parameter \( \beta = 0.0832 \) are shown in figure 3.3. The curves start at age two because we conditioned the age-specific survival analysis to individuals at least that age (see section 3.3.3). The probability \( \pi = 0.8925 \) of showing up in the colony conditional on being seen once is very high.

Birds are more likely to choose a partner that they know. The probability of choosing a certain partner increases with the number of times the pair has bred together in the past. In contrast, the previous breeding outcome with a partner does not influence the partner choice. The 95% CRI for the parameter \( \phi_{1} \), determining the effect of the accumulated number of fledglings produced in the past, includes zero and is therefore insignificant. Figure 3.4 shows the partner choice function.
when a bird has the choice between two different partners $A$ and $B$. In the left hand panel both partners are new at the beginning and the probability of choosing either partner is equal. The longer the bird bred with partner $B$ the more likely it becomes to pair with that partner again. In the right hand panel the bird has already bred two years with partner $A$ and partner $B$ would be new at the beginning. In this situation it is unlikely that the bird chooses partner $B$, but it becomes more likely the more often the bird breeds with partner $B$, with an equal probability of choosing either partner when the bird has spent two years with each of them in the past.

When accounting for the structure of the mechanism leading to the number of fledglings, the pair bond length is not important. The parameter $\omega_1$ for the influence of the number of previous years a pair has bred together is insignificant with a 95% CRI spanning from -0.0303 to 0.0323. The age of both partners is instead important for the breeding outcome. There is a period of improvement followed by senescence in both sexes. The age curves for males and female are very similar. Males reach their maximum reproduction only slightly later than females with $\omega_{3,1}$ being 11.2863 years (95% CRI: 8.8639-17.7662) compared to $\omega_{3,2}$ being 10.6958 years (95% CRI: 8.1607-15.1447). The steepness of the change in breeding performance with age $\omega_{2,s}$ is essentially the same for both sexes. The left hand panel of figure 3.5 shows the dependence of the number of fledglings produced on the interplay of the age functions for both sexes and individual quality for the model with the individual effects. Individual quality is defined in terms of reproduction and is given by the
Figure 3.4.: Function of the partner choice. In this example the bird has a choice between two partners, partner A and partner B. The graphs show the probability of breeding with partner B dependent on the number of times the bird has bred with that partner in the past, given that it has never bred with partner A (left) or has bred twice with partner A in the past (right). The grey area is the 95% predictive interval and the white line the predicted curve using the average estimated parameters.

individual effects. Figure 3.5 shows that it is best for the reproductive output to mate with a bird of intermediate age and of good quality.

Since the model with the individual effects has a much lower DIC value than the model without these effects there are clear differences between individuals. Comparing both models, the maxima of the age curves for both sexes lay at a later age when individual effects are not included (see the right hand panel of figure 3.5 for the model without the individual effects compared to the left hand panel for the model with the individual effects). The maximum for the males shifts upwards by $2.1467$ years to $\omega_{3,1} = 13.4330$ years (95% CRI: 10.1429-25.9333) and the maximum for the females by $1.5953$ year to $\omega_{3,2} = 12.2911$ years (95% CRI: 9.1497-19.0681). We plotted the individual traces for the marked individuals that were observed repeatedly and they showed convergence. We further compared the estimated individual effects to the average reproductive output during the observed period and they seem reasonable.

The probability of breeding a second and a third time in one breeding season is influenced by the breeding success on the previous occasion. Birds are much more likely to try to breed another time if the previous breeding occasion was unsuccessful. The probability of breeding a second time after an unsuccessful first breeding attempt is $0.2042$ (95% CRI: 0.1845-0.2246) and therefore not uncommon. In contrast,
the probability of breeding a second time after a successful first breeding attempt is 0.0145 (95% CRI: 0.0092-0.0216), which is very low. Breeding a third time is, in general, unlikely with a probability of 0.0260 (95% CRI: 0.0120-0.0512) after an unsuccessful second breeding occasion and 0.0083 (95% CRI: 0.0017-0.0289) after a successful second occasion. Figure 3.7 gives a graphical representation of the probabilities of breeding in an additional occasion dependent on the breeding success in the previous attempt.

3.5. Discussion

3.5.1. The Model

When accounting for the structure of the mechanism leading to the number of fledglings, the pair bond length is not important although the age of your partner is. Staying with the same partner does not consequently deliver a fitness advantage. Our approach is unusual because we simultaneously model multiple aspects of the breeding process, simultaneously consider females and males, while making
inferences about information missing from the data. Other models typically include only one level of the breeding process or consider the different processes separately (Coulson 1966; Mills 1973; Ollason and Dunnet 1978; Linden 1991; Williams and Rodwell 1992; Ens et al. 1993; Dhondt and Adriaensen 1994; Bradley et al. 1995; Briëd and Jouventin 1999; González-Solís et al. 1999a; Blondel et al. 2000; Mougin et al. 2000; Becker et al. 2001; Pyle et al. 2001; Streif and Rasa 2001; Green 2002; Saino et al. 2002; Saitou 2002; Heg et al. 2003; Moody et al. 2005; Pampus et al. 2005; van de Pol et al. 2006; Jeschke et al. 2007; Naves et al. 2006, 2007; Ludwigs and Becker 2007; Hatch and Westneat 2008; Arai et al. 2009; Nisbet and Dann 2009; Ismar et al. 2010). There are a few models that take more than one level into account but to our knowledge there are none that consider all the levels used in this study together. Cam et al. (2002) and Wintrebert et al. (2005) jointly modelled two steps, survival and breeding probability. Cam et al. (2002) used two individual random effects, one for each part of the model, and applied a Bayesian estimation procedure. Wintrebert et al. (2005) included the same individual random effect in both parts of the model. Wintrebert et al. (2005) used maximum likelihood methods and a Bayesian approach as alternative ways to estimate the unknown parameters, which both gave similar results. Knape et al. (2011) modelled survival and the number of fledglings for successful pairs simultaneously in a Bayesian framework. The link between the two parts is done through a correlation of the individual random effects in the survival and the breeding part. Sexes were not included simultaneously in the breeding part of the model. Kendall and Wittmann (2010) model the breeding process sequentially but do not include the pair formation process. They include a sequence of the probability of mating and nest building, the number of laid eggs, depredation of the eggs and the survival probability of the chicks to independence. Several breeding attempts per breeding season are possible in the model (Kendall and Wittmann 2010).

Males and females have only rarely been modelled simultaneously in other studies of divorce or mate retention before (Bradley et al. 1995; Blondel et al. 2000; Pyle et al. 2001; Saino et al. 2002; van de Pol et al. 2006). In most studies male and female breeding performance is analysed separately (Coulson 1966; Ens et al. 1993; Dhondt and Adriaensen 1994; González-Solís et al. 1999a; Streif and Rasa 2001; Green 2002; Balbontín et al. 2007; Ludwigs and Becker 2007; Arai et al. 2009; Nisbet and Dann 2009), only one sex is considered (Hatch and Westneat 2008), or it is not distinguished between sexes (Linden 1991; Williams and Rodwell 1992; Briëd and Jouventin 1999; Mougin et al. 2000; Becker et al. 2001; Black 2001; Saino et al. 2002; Saitou 2002; Heg et al. 2003; Moody et al. 2005; Naves et al. 2006, 2007;
Differences between individuals are often modelled in ecology and evolutionary biology by random individual intercepts, which requires the assumption of a certain distribution of these effects. We instead include individual identity as a fixed effect in our model and did not assume a specific distribution. This is commonly used in panel data analyses in human demography.

### 3.5.2. Our Results in Comparison to Previous Findings

**Divorce and Mate Retention**

In the literature there is support for different, sometimes even contradicting, divorce and re-mating hypotheses in wild bird populations. It is likely that several causes of divorce or mate retention act at the same time within a colony and that it is not one single hypothesis that explains every mate change (Heg et al. 2003).

Several studies found that pairs with breeding failure have a higher probability of divorce (black-legged kittiwake (Rissa tridactyla, Coulson 1966; Naves et al. 2007); Macaroni penguin (Eudyptes chrysolophus, Williams and Rodwell 1992); Cory’s shearwater (Calonectris diomedea borealis, Mougin et al. 2000)) or that those pairs that will subsequently divorce have a lower breeding success before separation than those that will re-mate with the same partner (red-billed gull (Larus novachollandiae scopulinus, Mills 1973); blue tit (Cyanistes caeruleus, Dhondt and Adriaensen 1994); common blackbird (Turdus merula, Streif and Rasa 2001)); common guillemot (Uria aalge, Moody et al. 2005)). There are other studies where breeding performance in the previous year did not influence the decision of divorce (Gentoo penguin (Pygoscelis papua, Williams and Rodwell 1992); white-chinned petrel (Procellaria aequinoctialis, Briëd and Jouventin 1999); blue tit (Blondel et al. 2000); barn swallow (Hirundo rustica, Saino et al. 2002; Arai et al. 2009); great tit (Parus major, Saitou 2002); Australasian gannet (Morus serrator, Ismar et al. 2010)). The result that breeding performance in the last breeding season did not have an effect on divorce probability was also found in several studies of the common tern (González-Solís et al. 1999a; Becker et al. 2001; Ludwigs and Becker 2007). Even though these studies only look at the year before divorce rather than all previous breeding attempts with a certain partner, the lack of an effect of previous breeding success for the mate decision accords with our finding.

There are contrasting results for the influence of pair bond duration on the divorce likelihood. Pair bond length had no effect on divorce probability in the white-chinned petrel (Briëd and Jouventin 1999) and the oystercatcher (Haematopus ostralegus,
van de Pol et al. 2006). In contrast, in the black-legged kittiwake (Naves et al. 2007) and the little penguin (Nisbet and Dann 2009) longer pair bonds make a mate change less likely, which is in accordance with our finding that the longer the pair bond length with a certain partner the more likely it is to choose that partner again. Our result is also supported by the previous finding that mate retention is generally high in the common tern (González-Solís et al. 1999a, c; Becker et al. 2001).

Linden (1991) showed in experiments with great tits that pairs with an artificially reduced clutch size or number of hatchlings had a higher divorce rate than the control group while pairs with an artificially increased clutch size or number of hatchlings had a lower divorce rate. They conclude that divorce is an active choice by which the individual escapes a disadvantageous breeding situation and ultimately increases its future reproductive output. This result is in lines with the ‘better option hypothesis’, which is also supported by several non-experimental studies. Birds initiating a divorce subsequently improved their breeding output (common guillemot (Moody et al. 2005)) or social status (oystercatcher (Heg et al. 2003)), while their previous mates did not (Heg et al. 2003; Moody et al. 2005). In a study of the common guillemot by Moody et al. (2005) the cause for divorce seemed to be insufficient parental behaviour on the part of the bird that was left. This bird fed the chicks with lower frequency than its partner before the divorce and in several cases caused egg or chick loss directly. This result suggests that the partner left behind is of low quality (Moody et al. 2005). In some studies females seemed to take the decision to divorce and only females improved their reproductive performance after divorce (blue tit (Dhondt and Adriaensen 1994; Blondel et al. 2000); common blackbird (Streif and Rasa 2001)). In the study by Ollason and Dunnet (1978) on Northern fulmars (Fulmarus glacialis) breeding success is generally lower before divorce than after. For the common tern the ‘better option hypothesis’ was not supported in a previous study by González-Solís et al. (1999a). In this study body mass was considered as a measure of individual quality. Body mass did not vary between pairs that will subsequently divorce and those that will remain together. For divorced pairs the individual body mass of the partners before the divorce was not different from the body mass for the partners in the newly formed pair (González-Solís et al. 1999a).

The better option that drives divorce is not always a partner of higher quality, the determinant can also be the option for a better quality territory (blue tit (Blondel et al. 2000)). There are studies that claim that the change in partner is only a consequence of a change in breeding site. Briéd and Jouventin (1999) found that in the white-chinned petrel divorce is mainly caused by disturbance of rats and a subse-
quent change in site. In the studies of black-legged kittiwakes by Naves et al. (2006) and Naves et al. (2007) the length of prior residence in nest site explained divorce probability better than the length of pair bond. In contrast Pyle et al. (2001) and Briëd et al. (2003) emphasised that mate fidelity is not a consequence of site fidelity. There was on one hand no relationship of longevity or lifetime reproductive success and site fidelity but there was on the other hand a correlation between the performance measures and mate fidelity (Cassin’s anklet (*Ptychoramphus aleuticus*, Pyle et al. 2001); comparative study of Procellariiformes (Briëd et al. 2003)). Therefore, site fidelity functions rather as a way to find the previous partner in these studies (Briëd et al. 2003). This is also supported in the study by Kim et al. (2007) on blue-footed boobies (*Sula nebouxii*) where divorce resulted in greater dispersal than faithfulness. Fairweather and Coulson (1995) even showed that for the black-legged kittiwake pairs remained together despite preventing the birds from returning to their previous nest sites. Thus, birds in a pair recognised each other and did not always simply return to the same nesting spot. Although fidelity to the nest site is higher in birds that re-mate with their previous partner than in divorced birds for the common tern (Becker et al. 2001), mate fidelity is also not just a consequence of site fidelity as there are birds that change the territory but re-mate with the same partner (González-Solís et al. 1999c).

Birds divorce more often in bad quality sites (blue tit (Blondel et al. 2000); oystercatcher (Heg et al. 2003; Bruinzeel 2007)). In accordance with the ‘forced divorce hypothesis’ in low quality sites birds are chased away by conspecifics (oystercatcher (Heg et al. 2003; Bruinzeel 2007)). Birds that are forced to leave a nest were found to have subsequently a lower breeding success, while the success of the remaining bird was unchanged and the success of the intruder increased (common guillemot (Jeschke et al. 2007)).

Saitou (2002) found that great tit pairs divorce most often in the year right after formation, which suggests support for the ‘errors of mate choice hypothesis’. The study by Freed (1987) of the tropical house wren (*Troglodytes aedon*) gives support for the ‘constraint hypothesis’. There was no opportunity for birds to leave the pair due to the limited number of available breeding spots but a larger number of potential breeders. Hatchwell et al. (2000) found support for the ‘inbreeding avoidance hypothesis’ in the observed strategies of dispersal in the long-tailed tit (*Aegithalos caudatus*). The study by Owen et al. (1988) on barnacle geese (*Brant leucopsis*) supports the ‘accidental loss hypothesis’ with a loss of the partner during migration. There are several studies that give support for the ‘musical chair hypothesis’. In pairs of black-legged kittiwakes that bred successfully in the previous breeding sea-
son asynchronous arrival in the breeding colony increased the probability of divorce (Naves et al. 2006). In barn swallows partners who divorce showed a higher asynchrony in arrival time compared to those that re-mate (Arai et al. 2009). For the common tern the arrival date in the breeding colony of both partners within a pair was correlated in the study by Ludwig and Becker (2008a). Several authors reported that asynchronous arrival of members of a pair in the colony led to a higher divorce rate (González-Solís et al. 1999a; Becker et al. 2001; Ludwigs and Becker 2007). But the synchrony in arrival times for members of a re-mated pair and formerly divorced birds with their new partner was not different (Ludwigs and Becker 2007). Although these studies give support for the ‘musical chair hypothesis’ for the common tern, González-Solís et al. (1999a) and Ludwig and Becker (2006) suggest that re-mating with the previous partner is the preferred option for the common terns because it avoids additional costs of mate search. They conclude that ‘synchrony in arrival facilitates pair bond maintenance rather than asynchrony promoting divorce, since divorce appears to be a side-effect of asynchrony and not an active decision’ (González-Solís et al. 1999a).

Several studies compare the breeding success of pairs newly formed after divorce and pairs that have bred together before. There are studies which found that birds re-mating with their partner from the previous year have a higher reproductive success than those that divorced and mated with a different partner (black-legged kittiwake (Coulson 1966); oystercatcher (Ens et al. 1993); blue tit (Blondel et al. 2000); house sparrow (Passer domesticus, Hatch and Westneat 2008); Australasian gannet (Ismar et al. 2010)). This might be explained by the ‘syndrome of the first-year bond’, the ‘pair-bond investment hypothesis’ or the ‘mate familiarity hypothesis’. Whereas other studies found no effect of divorce or re-mating with a familiar partner on breeding success (Gentoo and Macaroni penguin (Williams and Rodwell 1992); white-chinned petrel (Briëd and Jouventin 1999); barn swallow (Saino et al. 2002); blue tit and great tit (Pampus et al. 2005)), which is broadly in line with our research finding. For the common tern there are also contradicting results. When considering only the selected group of second-year breeders, Ludwigs and Becker (2007) found that after divorce the pairs had a higher reproductive output than faithful pairs, with a stronger effect for males. A comparison of the breeding success before and after divorce showed that birds improved their breeding success by divorce. The improvement could be a consequence of earlier arrival of the male. But as there were also repairings after an intermittent year of breeding with a different partner, the authors conclude that divorce is not an adaptive strategy but rather a result of asynchronous arrival (Ludwigs and Becker 2007). In contrast to the finding by
Ludwigs and Becker (2007), the breeding performance (breeding success versus no breeding success, number of fledglings, clutch size, egg volume) is not different for divorced and faithful birds in the study by González-Solís et al. (1999a). Further, the breeding performance did not change from before to after divorce (González-Solís et al. 1999a).

Van de Pol et al. (2006) manipulated pair bond length in oystercatchers by removing partners, which had an effect on the reproductive performance of the bird that was left behind and had to re-pair with a new partner. The authors conclude that divorce implies some costs. These could be costs in energy or timing. The baseline corticosterone level is lower in black-legged kittiwakes that remain with a previous partner compared to individuals that changed the partner. This is an indication of higher energetic demands for divorced birds (Angelier et al. 2007a). Several studies give support for an earlier timing of egg laying and breeding for birds that stay with a familiar partner compared to newly formed pairs (black-legged kittiwake (Coulson 1966); red-billed gull (Mills 1973); Northern fulmar (Ollason and Dunnet 1978); oystercatcher (Ens et al. 1993); great tit (Pampus et al. 2005); common guillemot (Jeschke et al. 2007)). This result has also previously been found in the common tern (González-Solís et al. 1999a). Laying date even advanced with increasing pair bond length in oystercatchers except for very long durations of over 10 years (van de Pol et al. 2006). An additional cost of divorce is the risk of not breeding at all in the coming breeding season (Ens et al. 1993). In the oystercatcher intermittent breeding is caused by divorce or death of the partner (Bruinzeel 2007), but widowed birds are more likely to breed in the season immediately after the loss of the partner than divorced birds (Ens et al. 1993). In the white-chinned petrel birds that divorced usually missed breeding times before they bred again (Briëd and Jouventin 1999). In the case of the common tern it was previously reported that some birds also did not breed after divorce. Those birds that bred after divorce arrived earlier in the colony than the birds that did not breed (González-Solís et al. 1999a). Another disadvantage of pairing with a new partner might be a lack of coordination within the pair. Males in newly established pairs of brown thornbills (*Acanthiza pusilla*) provisioned chicks with food at a lower frequency than males in pairs that had bred together before (Green 2002). In contrast, new pairs of house sparrows were not found to have any disadvantage in coordination of their presence at the nest (Hatch and Westneat 2008).

Some studies that test for an effect of pair bond length on breeding success give support for the ‘mate familiarity hypothesis’. This is the case for the fulmar (Ollason and Dunnet 1978). Pair bond duration was related to standardised annual fecundity
in barn swallows (Balbontín et al. 2007). Pair bond length and age influenced breeding performance in the little penguin. The maximum of the age function was 8 years. Breeding performance increased with pair bond length until at least 5 years. The relationships were weaker in females than in males (Nisbet and Dann 2009). Age of both partners and pair bond length also influenced whether a pair breeds successfully in short-tailed shearwaters (*Puffinus tenuirostris*), with a stronger effect of the pair bond length (Bradley et al. 1995). In Cassin’s auklets pair bond length has a positive effect on breeding success, when accounting for age and experience of the male and the female. Including individual as a categorical variable does not change the effect of the pair bond length. In contrast to our results the individual quality effect was not significant (Pyle et al. 2001). Pair bond length also influenced the reproductive output in the oystercatcher, where the effect of pair bond length had a quadratic form (van de Pol et al. 2006). Lifetime reproductive success increased with pair duration in barnacle geese (*Branta leucopsis*), when controlling for life span and time without a partner (Black 2001). But there are also studies in which breeding performance did not depend on pair bond length which concur with the results of our study. In another study of a selected group of only second-time breeding common terns the breeding success of faithful pairs did not change from the first to the second time of breeding together (Ludwigs and Becker 2007). Pair bond length as a continuous function was not studied before in the common tern. Briëd and Jouventin (1999) reported a relation of pair bond duration and breeding performance in white-chinned petrels. In a study of the common guillemot Jeschke et al. (2007) found that the breeding success of birds that bred until the end of the study with the same partner was not different from the success of those that changed the partner in between. In the house sparrow reproductive performance does not increase with pair bond duration when age is accounted for. Female age and quality rather seem to be the determinants of higher breeding success (Hatch and Westneat 2008). Pair bond length influenced the breeding success probability in the black-legged kittiwake, with an increase at the beginning and a decrease at long durations when Naves et al. (2007) analysed their complete dataset. The increasing effect vanished when groups of equal pair bond length were analysed separately or when only high-quality individuals were studied. This implies that the increase in breeding performance with pair bond length in the complete dataset is due to a selection effect, which highlights the importance of including the individual effects in our analyses.
Other Aspects of the Model

We trust the result of the unimportance of the pair bond length for the reproductive performance and the partner choice because our model is validated by the results for different sub-parts being consistent with previous analyses of the separate processes. The probability of showing up in the colony conditional on being seen once is very high. A high site fidelity after first breeding has been previously reported for the common tern by Becker et al. (2008b). Even though the probability of attending the colony is high, a few birds are not seen in certain years. As the recapture probability is close to one in the studied colony (Szostek and Becker 2011) these birds either skip a breeding season or visit a different colony. Breeding information for other colonies is very limited. The generally reported high site fidelity for the common tern after having bred once, makes it likely that the birds not attending the colony skip breeding.

The basic age pattern in breeding performance with improvement at early ages followed by senescence at old ages has previously been reported by several authors. Ezard et al. (2007) and Rebke et al. (2010) have shown an improvement in reproductive performance over a long period of life with senescence occurring only late in life. Performance was measured in both studies as relative reproduction, the number of fledglings an individual has in a particular year, minus the average number of fledglings in the entire breeding population in that year. Several other authors have shown age-specific changes in different aspects of reproductive performance for the common tern. Nisbet et al. (2002) found that individuals ≥18 years had larger clutch sizes, a higher number of fledglings and an earlier laying date than individuals of ages between 6-9. González-Solís et al. (2004) showed an advancing laying date until age 8, where it reached a plateau. Arrival date at the breeding colony advances with age too (Becker et al. 2008a). González-Solís et al. (2004) found a quadratic relationship for egg-volume with age-specific increases until stabilisation around age 5 and a slight decrease after age 13. The body mass of breeding common terns at arrival in the breeding colony, at incubation and during the time of chick rearing increases with age (Limmer and Becker 2007).

The quadratic age pattern in breeding performance implies that it is best for a common tern to mate with a partner of intermediate age and of good quality. It was shown in previous studies that birds of high quality were less likely to change their partner (Nisbet and Dann 2009). We found only small differences in the age pattern of reproductive performance for males and females, with a slightly later maximum for the males who also are older than females when breeding for the first time (Ludwigs and Becker 2002). The common tern is a socially monogamous bird.
Extra-pair paternity is rare (González-Solís et al. 2001). Clutton-Brock and Isvaran (2007) showed previously for different monogamous bird and mammal species that the pattern of reproductive success is similar in both sexes. Even though there are sex-specific differences in tasks, males and females both contribute to the incubation and chick provisioning in the common tern (Wiggins and Morris 1987; Becker and Ludwigs 2004). They are therefore likely to invest approximately equally into their young, which might explain the similar pattern for both sexes.

The comparison of the two nested models with and without the individual effects showed that there are differences in reproductive performance between individuals that cannot be neglected. The maximum of the age curves for males and females are shifted to higher ages when individual effects are not included. This is likely a sign for selective disappearance of individuals with lower reproductive success. However, the effect must be not very strong since we found no correlation of age at death and the individual effect for birds with transponder \( R^2 = 0.0264 \); figure 3.9). Rebke et al. (2010) previously showed that there is some selective disappearance in the studied common tern population, even though it plays a minor role compared to average individual change.

A low general probability of additional breeding attempts, with a higher probability after an unsuccessful breeding attempt and a lower probability after a previous successful attempt concurs with previous findings (González-Solís et al. 1999b; Becker et al. 2001; Becker and Zhang 2011). Very few pairs start a third breeding attempt. There were only 5 cases in the dataset, all after an unsuccessful first and second attempt. Having a successful previous breeding attempt likely does not give enough time within the constrained breeding season for another attempt. It was previously shown for the common tern that the arrival time and the time when a pair starts to breed within a season influenced the re-nesting probability (Wendeln et al. 2000; Becker and Zhang 2011). Further, it might not be worth spending the additional effort after a successful breeding event given that the probability of a successful breeding decreases at the end of the season (Wendeln 1997). In all 5 cases of a third breeding attempt it was unsuccessful. The birds that make more than one breeding attempt in a breeding season are likely of high quality (Wendeln et al. 2000; Becker and Zhang 2011).

3.5.3. Discussion of Model Details

In our model we used a Gompertz mortality function. This function was chosen because it is widely used and it is the simplest function applied in the code by Colcher and Clark (2011), with the fastest parameter convergence. We would like
to implement other functions like a Weibull, a Gompertz-Makeham, a Siler model or an exponential in the future and test which is the most appropriate for the common tern.

We assume that the decision of the partner is made by each individual independently and only on the basis of past experience with that partner. This assumption seems reasonable due to the insufficient amount of data to test more complicated hypotheses and in order not to complicate the complex model even more. Nevertheless other mechanisms might be possible. The partner choice might depend on the arrival time of both partners (González-Solís et al. 1999a; Becker et al. 2001; Ludwigs and Becker 2007). Neighbouring pairs might also be observed during breeding and their past experience might play a role in partner choice (Southern and Southern 1982). It might also be possible that partner choice depends more on cues expressing individual quality a partner shows in the current breeding season rather than on past experience with that partner (review in Bergeron et al. 2011).

An alternative function for the number of fledglings to the truncated Poisson with an upper level of 3 is a proportional odds model for a repeated ordinal response (Agresti 2002).

Partner changes within the season are rare. They only happen in years of high predation (González-Solís et al. 1999c; Ludwig and Becker 2008b). Due to this low incidence it is difficult to estimate parameters for changes within a season and we therefore ignored them.

We are aware that there are concerns with using DIC as a tool to assess fit for hierarchical models (see the discussion on the paper in Spiegelhalter et al. 2002) and we plan to implement reversible jump into a future extension of the model (Green 1995). However, we believe that a difference in DIC as large as 470 gives clear support for the model with individual effects.

3.5.4. Implications of the Main Findings

In contrast to most previous studies we focused on studying the effect of mate retention rather than testing the hypotheses of divorce (see 3.2). Consequently, we cannot distinguish between the different hypotheses for divorce if pairs split up, but we can shed light on some hypotheses of mate retention.

In our study individuals are more likely to choose a partner that they know, regardless of their previous breeding outcome, even though the pair bond length does not influence the reproductive performance of the pair and is thus no fitness advantage. The lack of an effect of pair bond duration on breeding success suggests that the ‘mate familiarity hypothesis’ does not explain the mating pattern in our
population. Birds seem to choose a partner in their first breeding season and tend to stick to that partner. The variation in age at first breeding for the common tern is only a few years and most individuals start reproducing at age 3 (Becker et al. 2001, 2008a). Especially in their first year of reproduction common terns mate with a partner of their own age (Ludwigs and Becker 2005), which is not necessarily their preferred choice but a question of availability of partners that are willing to mate with them (Ludwig and Becker 2008a). This explains the correlation of the ages of partners that are marked ($R^2 = 0.2704$; see the left hand panel of figure 3.8), which was also found previously in other studies of the common tern (González-Solís et al. 1999a; Nisbet et al. 2007; Ludwig and Becker 2008a). As reproductive performance first increases with age the ‘assured-age hypothesis’ seems to be an explanation for the influence of pair bond duration on partner choice.

In addition to the age, the quality of partners is also correlated within pairs ($R^2 = 0.297$; see the right hand panel of figure 3.8). We used pairs in which both partners are marked for this analysis of correlations within a pair and used only the first time a pair appeared if it was recorded more than once. This result suggests that birds choose their partner at the beginning of their reproductive career according to its state, e.g. reproductive quality, and tend to stay with that partner.
3.6. Supporting Information

Figure 3.6.: The shape of the link function (left) for the number of fledglings over age for males and the corresponding functions of the actual number of fledglings (right) depending on the quality of the individuals. Quality is here defined in terms of reproductive output. It enters the equation as a shift of the intercept in the link function, where the best quality corresponds to an individual effect of $u = 0$ and the worst quality to $u = -1$.

Figure 3.7.: The probability of breeding a second (left) and a third time (right) in one breeding season dependent on the breeding success on the previous occasion. The error bars represent 95% predictive intervals.
Figure 3.8.: Correlation of ages (left) and individual effects (right) of both partners within a pair when both partners are marked. When the same pair of partners appears more than once only the first time it appears is used for the analysis. The grey lines were estimated by linear regression. The coefficients of determination are $R^2 = 0.2704$ for the correlation of ages and $R^2 = 0.297$ for the correlation of qualities. The curves are estimated with an intercept of 3.0021 and slope of 0.4615 for the former and an intercept of 0.2857 and slope of 0.5297 for the latter.

Figure 3.9.: Correlation of age at death and the individual effect for birds with transponder. The grey line was estimated by linear regression. The coefficient of determination is $R^2 = 0.0264$, the intercept was estimated with 2.2472 and the slope with 8.3902.
CHAPTER 4

From the Price Equation to a Decomposition of Population Change

This chapter has been published as: Maren Rebke (2010). From the Price equation to a decomposition of population change. *Journal of Ornithology* DOI 10.1007/s10336-010-0589-6.
4.1. Summary

Heterogeneity is a pervasive problem for analyses in biology and related fields in which it is of interest to get information unbiased by a change in the composition of the population. I show here how a reformulation of the Price equation leads to a decomposition method to address this issue. The derived equation gives the exact contributions of the average change in the surviving individuals and the change due to selective disappearance to the aggregate population change.

4.2. Introduction

Heterogeneity within a population is a widespread problem in the analysis of individual patterns in biology and related fields. If selective mortality removes frail individuals (of lower quality) at an earlier age than less frail individuals (of higher quality) a selected group of individuals having a higher survival and reproductive success remains. This is known as selective disappearance or compositional change. Therefore, conclusions about average individual age-trajectories often cannot be drawn directly from observations made on the aggregate population (Vaupel et al. 1979; Vaupel and Yashin 1985; Cam and Monnat 2000). In this study, I show how the Price (1970) equation can be used as a starting point to derive a new simple equation to address the issue of heterogeneity in a population.

Price introduced his now well-known equation to analyse the change in the average value of a phenotypic trait from one generation to the next by decomposing it into a change due to selection by differential reproduction and the difference in the mean trait value between parents and their offspring as a result of imperfect transmission (Price 1970). His equation built on earlier analyses of selection by Robertson (1966) and Li (1967) with an additional component analysing the potential difference between parents and offspring.

Price (1972) later extended his equation originally developed for kin selection to study the change in gene frequency between two generations by deriving an equation for group selection. The Price equation has been recently extended and modified in different ways. Frank (1997) expanded the equation by including two additional steps: a description of characters by using multiple regression on predictors like alleles and a description of fitness by using multiple regression on these characters. Grafen (2000, 2002) introduced a multi-generational form and derived the link between the Price equation and optimisation programs. He further developed a version of the equation that uses inclusive fitness rather than individual fitness as the
measure that is targeted by selection by including social interactions (Grafen 2006). Heywood (2005) further partitioned the original Price equation into five distinct processes: a linear response to selection and four different ways of imperfections in transmissions of phenotypic traits from parents to their offspring. Coulson and Tuljapurkar (2008) extended the Price equation to an age-structured form to analyse retrospectively why the mean of a phenotypic trait changes over time. Rice (2008) introduced a stochastic version that treated offspring phenotype and fitness expressed as the number of descendants as random variables. Rather than studying known reproductive outcomes in hindsight, this version considered the possible result of future reproduction. Kerr and Godfrey-Smith (2009) generalised the Price equation in a way that allowed various connections between ancestors and descendents and added another term to the equation that addressed the effect of immigration or a mixture of uni- and biparental reproduction. In the field of demography, a similar formulation to the Price equation for continuous data was independently derived by Vaupel, and published in Vaupel and Canudas-Romo (2002), that separates a change in an overall average (over time or age) into compositional changes and the direct change occurring at the subgroup level.

Price (1970) himself noted that his equation can be applied to selection problems in general and not only genetic selection. Rice (2004) even calls it the ‘algebra of evolution’. Apart from evolutionary genetics and group selection, it has been applied in other fields: e.g. in community ecology (e.g. Collins and Gardner 2009), in epidemiology (Day and Gandon 2006, 2007), in economics (Andersen 2004), and even in linguistics (Jäger 2008).

The different extensions and applications show that the Price (1970) equation is useful in various contexts and still of great interest. It is well known and widely accepted and therefore a good basis for the development of other equations. In this study, I develop a new equation from the Price equation that exactly decomposes retrospectively observed change at the aggregate population level into relative contributions from change in the mean trait of the surviving individuals and selective disappearance. Surviving is here defined broadly as not leaving the study population. The population level refers to the aggregate overall average for all individuals in the population. Selective disappearance describes the change in the mean trait due to a change in the composition of the population caused by losses. Although my equation can be derived from the Price equation, it is a different and more intuitive formulation. The primary distinction is that in this equation I am studying the change in single individuals and not the difference between groups of individuals (e.g. parents and their offspring).
4.3. Methods

The Price (1970) equation gives the population change in gene frequency from one generation to the next $\Delta Q$ as:

$$\Delta Q = \text{Cov}(z, q) \frac{z}{\bar{z}} + \sum_{j=1}^{N} \Delta q_j \cdot z_j \frac{1}{N \cdot \bar{z}}.$$  \hspace{1cm} (4.1)

The summation is here taken over all parents $N$. The variable $z$ is the number of offspring with $z_j$ being the number for individual $j$ and $\bar{z}$ being the mean number for all parents. The difference $\Delta q_j$ gives the change in the frequency of the considered gene from parent to offspring for the parent individual $j$. The term Cov($z, q$) is the covariance of $z$ and $q$ in the population of the parents.

This equation can be redefined for our purpose to get the average change in a trait value at the population level $\Delta V(x)$. To avoid confusion, I use other notation:

$$\Delta V(x) = \text{Cov}(v, w) \frac{v}{\bar{w}(x)} + \sum_{j=1}^{N(x)} \left[ v(j, x + k) - v(j, x) \right] \cdot w(j, x) \frac{1}{N(x) \cdot \bar{w}(x)}.$$ \hspace{1cm} (4.2)

Here, $x$ is defined as age and $k$ is the respective length of the age interval considered. The term $w(j, x)$ denotes a weighting function that is either zero if the individual $j$ does not survive from age $x$ to age $x + k$ or 1 if the individual $j$ survives to age $x + k$. The average $\bar{w}(x)$ represents the mean of these weights for the population $N(x)$ at age $x$ and is in essence the proportion of individuals that survive from age $x$ to $x + k$. Therefore, $N(x) \cdot \bar{w}(x)$ is the number of survivors from age $x$ to age $x + k$. The variables $v(j, x)$ and $v(j, x + k)$ give the considered trait values for individual $j$ at age $x$ and at age $x + k$, respectively. Here, $v(j, x + k) = v(j, x)$ if $w(j, x) = 0$, otherwise $v(j, x + k)$ is the real trait value that individual $j$ has at age $x + k$. The term Cov($v, w$) denotes the covariance of the trait value at age $x$ and the weights indicating whether an individual will survive to the next age $x + k$ or not. The original Price equation looked at a parent and its offspring, I instead consider here the same individual at two different ages (Rice 2004).

Robertson (1966), Li (1967) and, later, Price (1970) independently pointed out that the difference in the mean of selected individuals and the mean of the entire population can be formulated as a scaled covariance. I use this equivalence here and reformulate the covariance in the Price equation in terms of differences, which gives
the following equation:

$$
\Delta V(x) = \frac{N(x)}{N(x) \cdot \bar{w}(x)} \sum_{j=1}^{N(x)} v(j, x) \cdot w(j, x) - \frac{N(x)}{N(x) \cdot \bar{w}(x)} \sum_{j=1}^{N(x)} v(j, x) \cdot \bar{w}(x)
$$

(4.3)

$$
+ \frac{N(x)}{N(x) \cdot \bar{w}(x)} \sum_{j=1}^{N(x)} [v(j, x + k) - v(j, x)] \cdot w(j, x)
$$

This can be simplified into:

$$
\Delta V(x) = \frac{N(x)}{N(x) \cdot \bar{w}(x)} \sum_{j=1}^{N(x)} v(j, x) \cdot w(j, x) - \frac{N(x)}{N(x) \cdot \bar{w}(x)} \sum_{j=1}^{N(x)} v(j, x) \cdot \bar{w}(x)
$$

(4.4)

$$
+ \frac{N(x)}{N(x) \cdot \bar{w}(x)} \sum_{j=1}^{N(x)} [v(j, x + k) - v(j, x)] \cdot w(j, x)
$$

As $[v(j, x + k) - v(j, x)] \cdot w(j, x) = 0$ for all individuals that do not survive from age $x$ to $x + k$ and $N(x) \cdot \bar{w}(x)$ gives the number of surviving individuals to the next age $x + k$, the last part of the equation can be simplified by reformulating it in terms of only the surviving individuals rather than all individuals:

$$
\Delta V(x) = \frac{N(x)}{N(x) \cdot \bar{w}(x)} \sum_{j=1}^{N(x)} v(j, x) \cdot w(j, x) - \frac{N(x)}{N(x) \cdot \bar{w}(x)} \sum_{j=1}^{N(x)} v(j, x) \cdot \bar{w}(x)
$$

$$
+ \frac{N(x)}{N(x) \cdot \bar{w}(x)} \sum_{j=1}^{N(x)} [v(j, x + k) - v(j, x)] \cdot w(j, x)
$$

(4.5)

where $l$ denotes each surviving individual from age $x$ to $x + k$. $M(x)$ is the total number of these surviving individuals. In other words, this term is the average change in the trait value of the individuals.

As $v(j, x + k) - v(j, x) = 0$ for individuals that do not survive from age $x$ to $x + k$, the first part of the equation can also be further reformulated by expressing the appropriate sections in terms of survivors rather than all individuals:

$$
\Delta V(x) = \frac{N(x)}{N(x) \cdot \bar{w}(x)} \sum_{j=1}^{N(x)} v(j, x) \cdot w(j, x) - \frac{N(x)}{N(x) \cdot \bar{w}(x)} \sum_{j=1}^{N(x)} v(j, x) \cdot \bar{w}(x)
$$

$$
+ \frac{N(x)}{N(x) \cdot \bar{w}(x)} \sum_{j=1}^{N(x)} [v(j, x + k) - v(j, x)] \cdot w(j, x)
$$

(4.6)

The same logic of considering means over different numbers of individuals (namely
all individuals or just the survivors) was previously used in the framework of the Price equation by Coulson and Tuljapurkar (2008). The first part of this equation is the mean trait value at age $x$ of those individuals that survive to age $x + k$ minus the mean trait value at age $x$ of all individuals alive at age $x$. This difference captures the change due to losses, or in other words the change due to selective disappearance.

The second part gives the average change of the individuals. Equation 4.6 is a simple but useful decomposition of population-level change. A graphical representation and summarising form of equation 4.6 is given in Rebke et al. (2010).

To show that the decomposition in equation 4.6 holds, it can be further simplified. The term $\frac{M(x)}{\sum_{l=1}^{M(x)} v(l,x)}$ cancels out and leaves:

$$\Delta V(x) = \frac{\sum_{i=1}^{M(x)} v(l, x + k)}{M(x)} - \frac{\sum_{j=1}^{N(x)} v(j, x)}{N(x)},$$

which is the definition of population-level change.

### 4.4. Example

For a better understanding of the equation, consider the following example of the change in the number of fledglings from age 5 to 6 years in the common tern (Sterna hirundo). This is an example adapted from Rebke et al. (2010).

I use a total of 1155 observations of reproduction (number of fledglings at a given age) by 612 individually-marked common terns of known age and known breeding status. Non-breeding birds enter the dataset with zero fledglings. Birds that are not detected in a given year but reappear in subsequent years are considered as non-breeding individuals. The data were collected in a common tern colony in the Banter See in Wilhelmshaven (German North Sea coast, 53°27’N, 08°07’E) between 1992 and 2007. They come from the philopatric part of the population, of which most are marked as native young. Common terns have been ringed in this colony since 1984 and all fledged chicks have been additionally marked with subcutaneously implanted transponders since 1992. The details of this ongoing long-term study are described in Becker et al. (2001). An automatic detection system ensured a recapture probability close to one (Becker et al. 2001, 2008b). This makes it possible to follow individuals over age.

The number of individuals at age 5 is $N(x) = 612$. Assuming no new birds enter the study at age 6, all birds alive at age 6 have to be survivors from age 5.
The number of surviving individuals from age 5 to age 6 is $M(x) = 543$. It thus automatically also represents the number of individuals at age 6. The change at the level of the population is given according to equation 4.7 by the difference between the mean number of fledglings of all individuals at age 6 and the same average at age 5:

$$\Delta V(5) = \frac{M(5)}{M(5)} - \frac{N(5)}{N(5)} = 310 - 258 \frac{543}{612} = 0.57 - 0.42 = 0.15.$$

We can follow the surviving individuals back and calculate their mean number of fledglings at age 5, which represents the first term of equation 4.6:

$$\frac{\sum_{l=1}^{M(5)} v(l, 5)}{M(5)} = 0.44.$$

The difference between this mean for the surviving individuals and the mean for all individuals at age 5 is the change due to selective disappearance and is represented by the first two terms of equation 4.6:

$$\frac{M(5)}{M(5)} - \frac{N(5)}{N(5)} = 0.44 - 0.44 = 0.02.$$

This difference consequently measures the compositional change due to differential survival.

The average difference for the survivors between age 6 and age 5 captures the average individual change and is represented by the last term of equation 4.6:

$$\frac{\sum_{l=1}^{M(5)} v(l, 6) - v(l, 5)}{M(5)} = 0.13.$$

Putting the different terms together into equation 4.6 gives:

$$\Delta V(5) = 0.02 + 0.13 = 0.15.$$

This result shows that, in my example, the average number of fledglings increases from age 5 to age 6 on the population level. Around 87% of this change is explained by average individual changes and only around 13% is due to selective disappearance of individuals.
4.5. Discussion and Conclusion

I showed here how a novel decomposition of population change into the exact components of change in the mean trait of the surviving individuals and change due to selective disappearance can be derived from the Price equation 4.1. The simplification leading to the new equation 4.6 is much more intuitive than the redefined form in equation 4.2 of the original Price equation 4.1. The two parts of equation 4.6 have a clear biological meaning: namely the average change at the individual level and the effect of a change in the composition of the population due to losses of certain individuals. While the covariance term in the Price equation giving the compositional change is non-intuitive, the difference term in my new equation describes this change in a straightforward way.

Other methods were previously proposed to account for heterogeneity within a population (e.g. McNamara and Houston 1996; Cam et al. 2002; Barbraud and Weimerskirch 2005; Beauplet et al. 2006; van de Pol and Verhulst 2006). Of these methods, the use of random individual intercepts in combination with some fixed effect controlling for the quality of individuals is most widely used (van de Pol and Verhulst 2006). Unlike the previous approaches which were not able to decompose the population change into the contributing components, this method can exactly quantify the relative contribution of various processes to observed age-specific population-level patterns.

The decomposition method requires longitudinal data with repeated measures of a trait. It is not possible to address an event directly that only happens once to an individual, for instance entry into an absorbing state such as death.

The approach introduced is of importance because heterogeneity within a population as a confounding factor is a pervasive challenge in biology and demography. In longitudinal studies of fertility, disability, migration and other demographic events over age, individual propensities and capabilities differ. The set of individuals in a study at some time are different from that set at a later time, because some individuals have left the study and others may have joined it. With this method, life history studies can account for compositional change. Rebke et al. (2010) further extended equation 4.6 to include an additional term capturing the selective appearance of individuals. This term is of importance if individuals enter the study at some later age due to immigration or postponed recruitment.

I presented here a short example of fertility change over age to illustrate the derived equation. Rebke et al. (2010) provide an extensive application. They applied equation 4.6 to the long-term individual-based reproductive data of common terns
to study the shape of reproductive performance over age. Not only can change over age be analysed with equation 4.6, it is possible to use any longitudinal setup that can be split into distinct intervals, for example a change over time or a change over parity. In general, the method is useful whenever the interest lies in changes at the individual level but the change observed at the population level is confounded by a compositional change of the population.
CHAPTER 5

Reproductive Improvement and Senescence in a Long-Lived Bird

This chapter has been published as: Maren Rebke, Tim Coulson, Peter H. Becker, and James W. Vaupel (2010). Reproductive improvement and senescence in a long-lived bird. Proceedings of the National Academy of Sciences 107, 7841-7846.

Co-author contributions: Peter H. Becker collected and provided the data. Tim Coulson and James W. Vaupel conceived the idea of applying the traditional Price equation, but I conceived the idea of simplifying it and came up with the new formulation. All co-authors commented on the manuscript, with considerable improvements made by Tim Coulson and James W. Vaupel on the structure and the wording of the manuscript.
5.1. Summary

Heterogeneity within a population is a pervasive challenge for studies of individual life histories. Population-level patterns in age-specific reproductive success can be broken down into relative contributions from selective disappearance, selective appearance of individuals into the study population and average change in performance for survivors (average ontogenetic development). In this study, we provide an exact decomposition. We apply our formula to data on the reproductive performance of a well-characterised population of common terns (Sterna hirundo). We show that improvements with age over most of adult life and senescence at old ages are primarily due to a genuine change in the mean among surviving individuals rather than selective disappearance or selective appearance of individuals. Average ontogenetic development accounts for approximately 87% of the overall age-specific population change.

5.2. Introduction

In the study of life histories it is important to capture the performance of individuals and changes in that performance over age. Changes in performance are often reported as changes in population averages, but individual heterogeneity can produce dynamics at the observed population level that are very different from dynamics at the individual level (Vaupel et al. 1979; Vaupel and Yashin 1985; Cam and Monnat 2000).

‘Population level’ refers to the observed averaged information over all individuals present at a given age regardless of whether they survive to the next age step. What we define as population change is the change observed directly in the aggregate field data and not the change in a hypothetical population of all individuals that were present at a given starting age $x$. ‘Individual level’ refers to the change over age for each surviving individual averaged over all these survivors. We call this term ‘average ontogenetic development’.

The interpretation of population-level patterns can be problematic due to ‘within-generation phenotypic selection’ (Endler 1986). This is a change in the composition of the population if selective mortality removes frail (of lower quality) individuals at an earlier age than those individuals that are less frail (of higher quality)(Vaupel et al. 1979; Vaupel and Yashin 1985), or if individuals that enter the population (here breeding population) as adults later than others through delayed recruitment or immigration (van de Pol and Verhulst 2006) have a different performance than the
resident population. We use the phrase ‘selective disappearance’ to denote change in the mean of a phenotypic trait due to mortality and ‘selective appearance’ to denote change due to new appearance of individuals. The term ‘compositional change’ is used to refer to the combination of both selective appearance and disappearance. We develop a method to exactly decompose population change in an age-specific phenotypic trait with repeated measures into the components: average ontogenetic development, selective disappearance and selective appearance.

5.3. An Approach to Decomposing Population Change

If there is no selective appearance, then population change \( P \) over age can be exactly additively decomposed into average ontogenetic development for the surviving individuals \( s \) plus compositional change due to selective disappearance \( d \):

\[
P = s + d, \tag{5.1}
\]

where

\[
P = V_{x+k} - V_x \tag{5.2}
\]

is the difference between the averages of a trait of all individuals at one age \( V_x \) and the next age \( V_{x+k} \) with \( k \) being the age interval,

\[
s = v_{x+k} - v_x \tag{5.3}
\]

gives the respective difference for the individuals that survive from age \( x \) to \( x+k \) at one age \( v_x \) and the next age \( v_{x+k} \), and

\[
d = v_x - V_x \tag{5.4}
\]

is the difference between the two means at the first age \( x \) (see section 5.5 and figure 5.1 for a specific case). The expected change due to selective disappearance is zero if the population is homogenous. Equation 5.1 holds because the survivors at the second age \( x+k \) constitute the entire population at that age if no individuals enter the study at age \( x+k \), so \( V_{x+k} = v_{x+k} \) and

\[
s + d = (v_{x+k} - v_x) + (v_x - V_x) \tag{5.5}
\]
Figure 5.1.: Graphical representation of decomposition of the aggregate change at the population level into average ontogenetic development and change due to selective disappearance. \( P \) is the aggregate change in the mean number of fledglings produced from one age to the next – in this specific case we use data between age 5 and age 6, \( s \) is the average ontogenetic development and \( d \) denotes the change due to selective disappearance. \( V_5 \) and \( V_6 \) are the averages of the number of fledglings at age 5 and 6 respectively and \( v_5 \) and \( v_6 \) are the corresponding averages for the survivors.

\[
\begin{align*}
  P &= v_{x+k} - v_x \quad \text{(5.6)} \\
  s &= V_{x+k} - V_x \quad \text{(5.7)} \\
  &= P \quad \text{(5.8)}
\end{align*}
\]

For the method to provide an exact description of change, all living individuals need to be measured in each year. The selective disappearance term can be reformulated into

\[
d = (1 - f)(v_x - v_x^-), \quad \text{(5.9)}
\]

where \( f \) gives the fraction surviving and \( v_x^- \) is the average of a characteristic for the individuals at age \( x \) that do not survive to age \( x + k \).

Equation 5.1 can be extended to account for new individuals entering the study at the second age \( x + k \) (see section 5.9.1 for the extension with selective appearance).

Our approach builds on previous decompositions of change, namely the Price equation and related decompositions (Price 1970; Vaupel and Canudas-Romo 2002;
Coulson and Tuljapurkar 2008). These equations express the term that captures selective disappearance in the form of a covariance (Price 1970; Vaupel and Canudas-Romo 2002; Coulson and Tuljapurkar 2008). We instead use a difference between the average trait of selected individuals (here survivors) and the mean of the entire population (Robertson 1966; Li 1967). The method we propose is simpler, more intuitive in our context and decomposes the change into parts with clear biological interpretation. The approach cannot estimate the latent frailty itself; it rather captures the effects of different processes leading to the aggregate observed population change in the mean of a phenotypic trait over an age step. We do not intend to separate out the total effects over the entire lifetime, we rather present an approach to decompose the population change at each age. Our approach describes the observed change retrospectively, following the same use of the Price equation recently developed by Coulson and Tuljapurkar (2008) and Ozgul et al. (2009). The formula is not intended to project forward in time – it is a way of decomposing observed change.

5.4. Application

In this study, we apply this approach to the problem of interpreting longitudinal data on age-specific reproductive performance for a long-lived bird species, the common tern (Sterna hirundo). It is widely believed that reproductive performance for birds and mammals increases at the earliest reproductive ages, reaches a plateau at intermediate ages and often begins to decline at older ages (Adams 1985; Clutton-Brock 1988; Newton 1989; Pugesek and Diem 1990; Weimerskirch 1992; Wooller et al. 1992; Forslund and Pärt 1995; Martin 1995; Cam et al. 2002; Laaksonen et al. 2002; McEligott et al. 2002; Mysterud et al. 2002; Reid et al. 2003; González-Solís et al. 2004; Rattiste 2004; Beauplet et al. 2006; Bowen et al. 2006; Catry et al. 2006; Nussey et al. 2006, 2009; van de Pol and Verhulst 2006; Weladji et al. 2006, 2008, 2010; Angelier et al. 2007b; Balbontín et al. 2007; Festa-Bianchet and King 2007; Hadley et al. 2007; Limmer and Becker 2007; Proffitt et al. 2007; Descamps et al. 2008; Jones et al. 2008; McCleery et al. 2008; Reed et al. 2008; Berman et al. 2009; Blas et al. 2009; Bouwhuis et al. 2009; Brown and Roth 2009). Only a few studies, however, have tried to correct for compositional change within the population (Cam et al. 2002; Rattiste 2004; Nussey et al. 2006; van de Pol and Verhulst 2006; Weladji et al. 2006; Balbontín et al. 2007; McCleery et al. 2008; Bouwhuis et al. 2009; Brown and Roth 2009; Nussey et al. 2009; Weladji et al. 2010). Different authors have speculated about the importance of selective disappearance and individual changes...
but they have not been able to quantify the contributions (e.g. Pugesek and Diem 1990; Weimerskirch 1992; Wooller et al. 1992; Gaillard et al. 2003; Weladji et al. 2006, 2008, 2010; Balbontín et al. 2007; Hadley et al. 2007; Limmer and Becker 2007; McCleery et al. 2008; Bouwhuis et al. 2009; Brown and Roth 2009; Steenhof and Heath 2009). Such quantification can be carried out using the approach that we are proposing here.

We use our approach to decompose population change of individually marked common terns of known age between age 3 and 21 years into contributing factors. Data on reproduction were collected from a long-term study between 1992 and 2007 (see section 5.8). Over this period, recapture rates have been exceptionally high in the considered age groups, being almost 100% thanks to marking with passive implanted transponder (PIT) tags and an automatic detection system (Becker et al. 2008b). We only have information for the breeding season. Data come from the philopatric part of the breeding population, most of which were marked as hatchlings. Immigrants are not marked. Selective appearances are therefore birds that are born in the colony that returned to breed at an age greater than the modal age of 3 (i.e. delayed recruitment) and those >3 years old that were present at the beginning of the study. The common tern is a monogamous long-distance migrant. Breeding pairs typically remain together for many years. Adults, on average, fledge one chick per year (range 0-3). Most individuals start breeding at 3 years (Becker et al. 2001, 2008a). Maximum longevity from our study is 21. The average life expectancy at age 3, 6.8 years, is much shorter than the length of the study.

We consider two measures of performance. First, we decompose change in the absolute number of fledglings $y_x$ produced between age $x$ and $x+1$. Because environmental variation generates temporal fluctuation in reproductive success in this population (Becker 1998), we next decompose change in relative (or residual) reproduction (as, e.g. in refs. Ezard et al. (2007); Limmer and Becker (2007); Becker et al. (2008b); Limmer and Becker (2010)). Relative reproduction $r_{ixt}$ gives the reproduction of an individual in relation to the averaged reproduction of the population in a given year. More specifically, relative reproduction is defined as the number of fledglings $y_{ixt}$ produced by an individual $i$ at age $x$ in a specific year $t$ minus the average fledgling rate in that year $t$ for all known-age birds $j$ with age $z$ between 3 and 21:

$$r_{ixt} = y_{ixt} - \frac{\sum_{z=3}^{21} \sum_{j=1}^{N_{zt}} y_{jzt}}{\sum_{z=3}^{21} N_{zt}}.$$  \hspace{1cm} (5.10)
The number of all known-age birds at age \( z \) and year \( t \) is \( N_{zt} \).

We then examine differences between the sexes by decomposing change separately for females and males. Next, we identify reasons for the age-specific patterns we observe by calculating decompositions and fitting various mixed models (Fitzmaurice et al. 2004) to each individual’s changes in reproductive performance at each age (see section 5.9.2) with age replaced by (i) the past number of breeding attempts, (ii) the past number of successful breeding events, and (iii) the length of the pair bond.

5.5. Example

To understand our approach, consider the following example of the change in the number of fledglings from age 5-6 years (figure 5.1). The change at the level of the population is given by the difference between mean number of fledglings of all individuals at age 6 (0.57) and the same average at age 5 (0.42). The difference between these values is \( P = 0.57 - 0.42 = 0.15 \). Assuming no new birds enter the study at age 6, all birds alive at age 6 have to be survivors from age 5. We can follow these surviving individuals back and calculate their mean number of fledglings at age 5 (0.44). The difference for the survivors is \( s = 0.57 - 0.44 = 0.13 \): this quantity captures the change in the mean trait among survivors i.e. average ontogenetic development. The difference between the two means at age 5 (\( c = 0.44 - 0.42 = 0.02 \)) measures the change in the composition of the population due to differential survival.

5.6. Results

At the population level, relative reproduction increases with age until approximately age 14 before decreasing (figure 5.2). Results for both absolute and relative change are similar. Furthermore, there is no linear relationship between the mean age in the population and year (\( r^2 = 0.01145 \), 14 df, \( P = 0.6933 \)); such a relationship might have removed part of the age effect when yearly means in reproductive success are corrected for relative reproduction. We report results for relative reproduction (see figure 5.5 for absolute reproduction). By applying our approach we calculate that, averaged over all individuals and all ages, 86.6% of the age-related change is attributable to average change in performance among surviving individuals and 14.7% to selective disappearance (figure 5.3A). Selective appearance at ages after 3 years does not play an important role, with a contribution of -1.3% (figure 5.3A) and is subsequently not reported. Each average relative reproduction term for the
average over all ages entering the equation \((V_x = V_M, v_x = v_M, V_{x+k} = V_{M+1}, v_{x+k} = v_{M+1})\) is calculated by summing over all individuals \(i\) or all surviving individuals \(l\) in all relevant ages \(z\) and dividing by the respective number of individuals (either \(N\) for all individuals or \(S\) for the survivors). For the averages \(V_M\) and \(v_M\) individuals in all ages but the last age enter the calculation:

\[
V_M = \frac{\sum_{z=3}^{20} \sum_{i=1}^{N_z} r_{izt}}{\sum_{z=3}^{20} N_z},
\]  

(5.11)
Reproductive Improvement and Senescence in a Long-Lived Bird

\[ v_M = \frac{\sum_{z=3}^{20} \sum_{l=1}^{S_z} r_{zt}}{\sum_{z=3}^{20} S_z}. \] (5.12)

\[ V_{M+1} \text{ and } v_{M+1} \text{ are averaged over individuals in all ages except the first considered age:} \]

\[ V_{M+1} = \frac{\sum_{z=4}^{21} N_z \sum_{i=1}^{r_{izt}}}{\sum_{z=4}^{21} N_z}, \] (5.13)

\[ v_{M+1} = \frac{\sum_{z=4}^{21} S_z \sum_{i=1}^{r_{izt}}}{\sum_{z=4}^{21} S_z}. \] (5.14)

The changes due to the different processes are given as a percentage of the population change with a negative sign representing a contribution that reduces change. Figure 5.2 summarises the reproductive patterns over age produced by each of the three factors contributing to population change.

The contribution to population change of average ontogenetic development varies with age. Visual examination of the data (figure 5.2 and figure 5.6) reveals that up to age 14, average individuals improve in reproductive performance, whereas their performance mean declines beyond this age. We consequently separated individuals into two groups: those at least 14 (when only 20% of the cohort at age 3 is still alive) and those younger than age 14. At younger ages, the change at the population level in relative reproduction is positive and is mainly an effect of an improvement in average individual performance; in comparison, selective disappearance has a small influence (figure 5.3B). In contrast, at older ages, average individual relative reproduction decreases with age (figure 5.3B), which shows that there is senescence at old age. However, this evidence needs to be treated with caution because of high variability around the average individual change contributions at advanced ages.

Our results provide evidence that the age-specific pattern of reproduction in the common tern is mainly due to changes in the mean among survivors. That average reproductive ability improves over an extended period of their life also holds for females (figure 5.7) and males (figure 5.8) when considered separately, and when relative reproduction is replaced with the number of fledglings produced (figure 5.5). Males on average seem to improve to a greater extent at young and middle ages and decline more slowly at advanced ages than females, with selective disappearance
Figure 5.3.: Change in relative reproduction from age $x$ to $x + 1$ averaged over all individuals and all ages. (A) Decomposition of population change into average ontogenetic development, selective disappearance and selective appearance in percent (959 individuals with 4307 observations). (B) Change in relative reproduction for the age groups 3-13 and 14-20, respectively, at the population level (dark blue bars), and the contributions to this change due to average ontogenetic development (light blue bars) and selective disappearance (orange bars; 955 individuals with 4303 observations). In the old age group the average individual decrease in reproductive performance is also apparent at the population level but to a smaller extent because of the slight tendency of less reproductive individuals to suffer higher mortality. The error bars represent 95% confidence intervals.

Breeding performance not only increases with age but also with past breeding experience, the number of past breeding successes, and mate familiarity measured by the length of pair bond (table 5.1). Changes in individual reproductive performance among surviving individuals are best explained by the number of years an individual produced fledglings, both for males and females. This is mainly attributable to a large improvement when having the first successful reproduction (table 5.2).
5.7. Discussion

The problems of inferring average individual-level processes from population-level patterns when there is selective disappearance are well known. Several researchers tacked onto population-level studies further analyses to elucidate individual-level change (refs. Weimerskirch 1992; Laaksonen et al. 2002; Reid et al. 2003; see section 5.9.3 for details). In addition, an array of methods have been derived to correct for selective disappearance or selective appearance (refs. McNamara and Houston 1996; Cam et al. 2002; Reid et al. 2003; Rattiste 2004; Barbraud and Weimerskirch 2005; Beauplet et al. 2006; Nussey et al. 2006, 2009; van de Pol and Verhulst 2006; Weladji et al. 2006, 2010; Balbontín et al. 2007; McCleery et al. 2008; Aubry et al. 2009; Bouwhuis et al. 2009; Brown and Roth 2009; Moyes et al. 2009; see section 5.9.3 for details).

The use of a random individual intercept is the only method proposed so far that can account for unobserved heterogeneity within the population. The combination with some fixed effect controlling for the quality of individuals is currently the approach most widely used to correct for selective disappearance or selective appearance (Rattiste 2004; Nussey et al. 2006, 2009; van de Pol and Verhulst 2006; Weladji et al. 2006, 2010; Balbontín et al. 2007; McCleery et al. 2008; Aubry et al. 2009; Bouwhuis et al. 2009; Brown and Roth 2009; Moyes et al. 2009). The method assumes that, with appropriate fixed effects, individual random intercepts can be interpreted as a latent variable often termed ‘frailty’ and that corrected estimates are no longer biased by a compositional change and can be interpreted as the expectation of individual performance (Vaupel et al. 1979; van de Pol and Verhulst 2006; Nussey et al. 2008). The choice of the fixed effect can affect the results (Moyes et al. 2009). Our approach does not require the estimation of, and correction for, latent variables. Our simple approach provides a powerful way of exactly quantifying contributions of various processes to observed age-specific population-level patterns. Previous approaches were not able to decompose the population change exactly into the contributing components.

In addition to being a confounding factor in longitudinal studies of fertility changes over age, heterogeneity is a pervasive challenge in biology and demography. Therefore, our formula is of broad interest in a wide range of different applications to decompose population change into average change of the surviving individuals and change in the population’s composition (Vaupel et al. 1979; Vaupel and Yashin 1985; Vaupel and Canudas-Romo 2002). The approach is not limited to quantitative traits; the formula holds for means in general. It is therefore also possible to
additively decompose categorical variables, in which case the means in the equation are substituted by proportions. For the specific case of a binary variable, this is the proportion of successes – i.e. the mean number of successes is considered.

If the contribution of selective disappearance and selective appearance to population patterns of age-specific reproductive performance is small, inferences from previous analyses of the study population that do not account for these biases are broadly correct (Nisbet et al. 2002; González-Solís et al. 2004; Ezard et al. 2007; Limmer and Becker 2007; Becker et al. 2008a). In cases where selective disappearance and selective appearance – including immigration – are influential, a failure to correct for them could substantially bias results. In our study, averaged over all reproductive ages, 87% of the population-level change is explained by average ontogenetic development. Although population-level change and average ontogenetic development show a similar pattern over age, the difference between them is not consistent (figure 5.2 and figure 5.6). However, the single-age patterns need to be treated with caution because of high variability in the data.

Our findings show that average individual improvement characterises a long period of life and senescence appears only to become important at old ages in the common tern. Only 20% of all mature birds reached the age when senescence may become important. As a result, the reproductive life of most individuals is characterised by an improvement in individual performance with age.

Are these results likely to be general? There are only a few comprehensive analyses that appropriately account for unobserved heterogeneity. For reproductive improvement in early life these studies are: reindeer (Rangifer tarandus, Weladji et al. 2006, 2010), barn swallows (Hirundo rustica, Balbontín et al. 2007), common gulls (Larus canus, Rattiste 2004), mute swans (Cygnus olor, McCleery et al. 2008), wood thrushes (Hylocichla mustelina, Brown and Roth 2009), the laying date in oystercatchers (Haematopus ostralegus, van de Pol and Verhulst 2006) and the breeding probability in black-legged kittiwakes (Rissa tridactyla, Cam et al. 2002). We found only two studies that do not show any improvement in certain reproductive performance measures when accounting for compositional change although there was an apparent improvement when not accounting for the bias. In the black-legged kittiwakes, the entire apparent improvement in breeding success (success or failure) was due to selective appearance and selective disappearance (Aubry et al. 2009). The apparent improvement in egg volume for the oystercatcher in the study of van de Pol and Verhulst (2006) was entirely explained by selective appearance of individuals laying larger eggs at older ages. Our results thus seem typical of findings from other vertebrates suggesting that reproductive improvement may be a widespread
Reproductive Improvement and Senescence in a Long-Lived Bird

phenomenon.

In some of the studies listed in the previous paragraph and in some additional studies accounting for unobserved heterogeneity, reproductive senescence was previously found for red deer \((Cervus elaphus, \) Nussey et al. 2006, 2009), barn swallows \((Balbontín et al. 2007)\), black-legged kittiwakes \((Rattiste 2004)\), mute swans \((Mc-Cleery et al. 2008)\), wood thrushes \((Brown and Roth 2009)\), great tits \((Parus major, \) Bouwhuis et al. 2009) and reindeer \((Weladji et al. 2010)\). Note that none of these studies of reproductive improvement and senescence decompose change at the population level into average individual change and compositional change.

In our study, the age-specific patterns in average improvement among survivors in reproductive performance between age 3 and 14 may be due to a gain in experience \((Forslund and Larsson 1992; \) Cichoń 2003; Weladji et al. 2006, 2008) – individuals may get better at raising fledglings, possibly through improved foraging ability \((Cichoń 2003; \) Limmer and Becker 2007, 2009). Alternatively, older birds may be less likely to reduce parental care when stressed \((Heidinger et al. 2006; \) Angelier et al. 2007b). For long-lived species it is unlikely that increasing reproductive effort because of a small probability of successful future reproduction is responsible for increasing fertility early in life \((Forslund and Larsson 1992; \) Forslund and Pärt 1995).

What do our results tell us about evolution? Different models have been developed to estimate optimal life history strategies by optimal allocation of resources among maintenance, growth, and reproduction, starting with the work of Gadgil and Bossert (1970). Increasing fecundity with age is typically predicted in indeterminate growers \((Kitahara et al. 1987; \) Kozłowski 1991; Reznick et al. 2002). Birds, however, are determinate growers and most species reach their final size before they start to fly – long before the onset of reproduction \((Charnov 1993)\). Only a few models address reproductive performance rather than reproductive effort for determinate growers \((Pianka and Parker 1975; \) Cichoń 2001; Kindlmann et al. 2001; Baudisch 2008). Baudisch (2008) shows that, depending on the returns of investment in reproduction or maintenance and growth up to maturity, different reproductive strategies are optimal, including improvement at all ages, improvement followed by senescence, and senescence at all ages. The other models \((Pianka and Parker 1975; \) Cichoń 2001; Kindlmann et al. 2001) predict a triangular pattern in reproductive performance over age.

In the classic evolutionary theory of ageing, performance is generally predicted to decline because the strength of selection against deleterious mutations decreases with age after maturity. Mortality reduces cohort size even if there is only extrinsic
mortality. This leads to a decrease with age in the net effect of a mutation for the 
population (Medawar 1952; Charlesworth 1973), implying that individuals should 
suffer senescence after maturity (Medawar 1952; Williams 1957; Hamilton 1966; 
Kirkwood 1977). As Williams (1957) put it, ‘the time of reproductive maturation 
should mark the onset of senescence’, at least for species reaching their final size at 
or before maturity. This corresponds to the predictions of Hamilton (1966). Even 
though these theories are generally discussed with a focus on mortality, the same 
reasoning should hold for reproductive performance (Hamilton 1966). According 
to the equations of Hamilton (1966) and Emlen (1970), the force of selection on 
age-specific mutations that act on fertility declines with age, which implies that fer-
tility should decrease over the adult lifespan (Partridge 1989; Baudisch 2005). Even 
though Emlen (1970) unconvincingly argues later in his article that fertility should 
first increase after maturity, he arrives at the direct conclusion of his equation that 
‘traits increasing fecundity will be pushed to earlier and earlier ages until stopped 
by opposing factors’. In ageing research, this classic viewpoint based on a declining 
selection pressure with age is still widely accepted (Flatt and Promislow 2007).

Application of our method to common terns provides strong support for the pre-
dictions of life history theory with a triangular shape in reproductive performance 
over age. The classic view of senescent declines from maturity is not observed. Our 
results add to the view that reality is more complicated than the classic evolutionary 
theory of ageing suggests. We conclude that, in common terns, reproductive senes-
cence occurs, but only after a prolonged period of improved age-related performance 
beyond the age at maturity.

5.8. Methods

5.8.1. Study Details

We use a total of 4307 observations of reproduction of 959 individually marked 
common terns of known age between age 3 and 21 years. The data were collected 
in a common tern colony in the Banter See in Wilhelmshaven (German North Sea 
coast, 53°27'N, 08°07'E). The Banter See colony is monospecific. The common terns 
nest there on six artificial islands, which are arranged in a distance of 0.9 m to the 
neighboring island, each 10.7×4.6 m and surrounded by a 60-cm wall.

Common terns have been ringed since 1980 at other places in Wilhelmshaven 
and when the colony started to establish in 1984, at the Banter See. Since 1992, 
all fledged chicks have been ringed and marked with subcutaneously implanted
transponders (TROVAN ID 100; TROVAN). Antennae on 44 elevated resting platforms read the individual specific alphanumeric code automatically every 5-10 sec for each individual within a distance ≤11 cm. Additional antennae were placed temporarily for 1-2 days during the breeding season around every incubated clutch. The details of this automatic detection system are described by Becker et al. (2001, 2008b). Only chicks and 101 breeders have been PIT-tagged since 1992. Therefore, there are birds at the colony that do not have an implanted transponder; they are mainly immigrants. As ringing commenced in 1980, 12 years before PIT-tagging started, the ages of most residents in the population were known at the beginning of the 16 years of data we use with a wide age range present in all years.

The colony is checked for new eggs, hatched chicks, and fledged chicks every 2-3 days during the remainder of the breeding season. A chick was defined as fledged when it reached an age of 18 days and is not found dead during the rest of the breeding season. Further details of this ongoing long-term study, restricted to the philopatric segment of the population, are described by Becker et al. (2001).

Because we are interested in adult life, the starting age in our analysis is 3 years. The last age considered is 21, the age of the oldest bird. Reproduction occurs in discrete birth pulses during a 3-month period every year. Most individuals have only one clutch per breeding season with a maximum of three eggs. However, in the case of complete clutch failure, some pairs lay additional eggs. Here, we consider the overall number of fledglings per year, which exceeded 3 in only one case. We excluded birds with unknown breeding status and unknown age. Nonbreeding birds entered the dataset with zero fledglings. We further assumed that birds that were not detected in a given year but reappeared in following years did not breed in that particular year. However, this number was small and did not influence the results. To study the sensitivity of the results to changes in the data, we used jackknifing to estimate 95% confidence intervals.

5.9. Supporting Information

5.9.1. Supporting Equation

Equation 5.1 can be extended to account for new individuals entering the study at the second age. Kerr and Godfrey-Smith (2009) previously incorporated an additional covariance term into the Price equation that illustrates the effect of immigration for change in an average trait value from one generation to the next. We instead use a difference to capture that term. The extended equation implies that
the mean of the considered characteristic at the second age for all individuals $V_{x+k}$ is no longer equal to the mean for the survivors $v_{x+k}$, unless the mean of the considered characteristic of the newly appearing individuals is not different from the average of this characteristic of the individuals that were in the study before. Therefore the term

$$a = V_{x+k} - v_{x+k},$$  \hspace{1cm} (5.15)

which represents the selective appearance of new individuals, has to be added to equation 5.15:

$$P = s + d + a.$$  \hspace{1cm} (5.16)

This extended equation is again exact if all living individuals are measured.

The mean of the characteristic for the newly entering individuals is a function of $v_{x+k}$ and $a$, as well as the number of survivors and the number of newly entering individuals. If the mean for the newly entering individuals at the second age is higher than the mean for the individuals that were already in the study before (the survivors), then the aggregate mean $V_{x+k}$ is higher than the mean for the survivors $v_{x+k}$ and $a$ is positive (figure 5.4A). In the opposite case the term $a$ is negative (figure 5.4B).

### 5.9.2. Mixed Models

Our mixed models (Fitzmaurice et al. 2004) all include individual identity as a random effect. Age, the number of past breeding events, the number of past successful breeding attempts, or the pair bond length is fitted as a linear fixed effect and/or quadratic. We also modelled these factors by categorical groupings. The models are fitted to data on individual change in relative reproduction from one age $r_{ixt}$ to the next $r_{i(x+1)(t+1)}$:

$$\Delta r_{ixt} = r_{ixt} - r_{i(x+1)(t+1)}$$  \hspace{1cm} (5.17)

We only consider individuals for this analysis when we have information on age, the number of past breeding attempts, the number of past successful breeding events, and the length the bird bred together with its partner. We consider males and females separately. The mixed models are fitted via maximum likelihood (ML) in R version 2.7.0 by using the function `lmer` in package `lme4`. 
5.9.3. Supporting Discussion

Analyses beyond population-level analyses to elucidate individual-level change include: a separate longitudinal analysis for a restricted sample of individuals that survive to old age (Laaksonen et al. 2002; Reid et al. 2003); examination of correlation in improvements in survival and breeding success (Weimerskirch 1992); and a test for a relationship between breeding lifespan and reproductive performance (Reid et al. 2003).

Methods to correct for selective disappearance or selective appearance include: the arbitrary classification of individuals into groups (Reid et al. 2003) and inclusion of individual covariates in the model to capture observed heterogeneity (Beauplet et al. 2006); the construction of multistate models with states designed to account for the different qualities of individuals (McNamara and Houston 1996; Barbraud and Weimerskirch 2005); the joint modelling of survival and reproduction with individual identity included as a random effect (Cam et al. 2002); and correction for individual quality by including observed individual quality measures like longevity or age at first reproduction as fixed effects, and individual identity as a random effect (usually random intercept and no random slope) in linear, generalised linear mixed models (reference for model types: Fitzmaurice et al. 2004; references for applications: Rattiste 2004; Nuusy et al. 2006; van de Pol and Verhulst 2006; Weladji et al. 2006; Balbontín et al. 2007; McCleery et al. 2008; Bouwhuis et al. 2009; Brown and Roth 2009; Moyes et al. 2009; Nuusy et al. 2009; Weladji et al. 2010), and generalised additive mixed models (reference for model type: Wood 2006; reference for application: Aubry et al. 2009).
Figure 5.4.: Graphical representation of Equation 5.16 to decompose the aggregate change at the population level into average ontogenetic development, change due to selective disappearance, and change due to selective appearance when newly entering individuals have a higher mean than the surviving individuals already in the study (A) and newly entering individuals have a lower mean than the surviving individuals (B). $P$ is the aggregate change in the considered characteristic from age $x$ to the age $x+k$, $s$ is the average ontogenetic development, $d$ denotes the change due to selective disappearance, and $a$ the change due to selective appearance. $V_x$ and $V_{x+k}$ are the averages of the characteristic on the observed population level at age $x$ and $x+k$, respectively, and $v_x$ and $v_{x+k}$ are the corresponding averages for the survivors.
Figure 5.5.: Change in number of fledglings from age $x$ to $x+1$ averaged over all individuals and all ages. (A) Decomposition of population change into average ontogenetic development, selective disappearance and selective appearance in percent (959 individuals with 4307 observations). The proportional change is equivalent to an absolute mean change from one age to the next of 0.091 fledglings at the population level, a positive average ontogenetic development of 0.065; a change of 0.015 is due to selective disappearance and 0.01 to selective appearance. (B) Change in number of fledglings for the age groups 3-13 and 14-20, respectively, at the population level (dark blue bars), and the contributions to this change due to average ontogenetic development (light blue bars) and selective disappearance (orange bars; 955 individuals with 4303 observations). The error bars represent 95% confidence intervals.
Figure 5.6.: Change in relative reproduction from age $x$ to $x + 1$ at the population level (dark blue bars), and the decomposition of this change into average ontogenetic development (light blue bars) and selective disappearance (orange bars; 955 individuals with 4303 observations). The error bars represent 95% confidence intervals.
Figure 5.7.: Change in relative reproduction from age $x$ to $x+1$ averaged over all females and all ages. (A) Decomposition of population change into average ontogenetic development, selective disappearance, and selective appearance in percent (411 individuals with 1893 observations). (B) Change in relative reproduction for the age groups 3-13 and 14-20, respectively, at the population level (dark blue bars) and the contributions to this change due to average ontogenetic development (light blue bars) and selective disappearance (orange bars; 408 individuals with 1890 observations). The error bars represent 95% confidence intervals.
Figure 5.8.: Change in relative reproduction from age $x$ to $x+1$ averaged over all males and all ages. (A) Decomposition of population change into average ontogenetic development, selective disappearance and selective appearance in percent (468 individuals with 2123 observations). The average change among survivors contribution can be higher than 100% if the changes due to selective disappearance and/or selective appearance go in the opposite direction. (B) Change in relative reproduction for the age groups 3-13 and 14-20, respectively, at the population level (dark blue bars) and the contributions to this change due to average ontogenetic development (light blue bars) and selective disappearance (orange bars; 467 individuals with 2122 observations). The error bars represent 95% confidence intervals.

Table 5.1.: Decomposition of population change in relative reproduction from event $x$ to $x+1$ in percent

<table>
<thead>
<tr>
<th>Event</th>
<th>Average ontogenetic development</th>
<th>Selective disappearance</th>
<th>Sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of past breeding attempts</td>
<td>96.8% (± 11.6%)*</td>
<td>3.2% (± 11.6%)*</td>
<td>701 individuals (3078 observations)</td>
</tr>
<tr>
<td>Number of past successful breeding attempts</td>
<td>88.8% (± 16.4%)*</td>
<td>11.2% (± 16.4%)*</td>
<td>489 individuals (1454 observations)</td>
</tr>
<tr>
<td>Pair bond length in years</td>
<td>90.1% (± 36.9%)*</td>
<td>9.9% (± 36.9%)*</td>
<td>288 pairs (834 observations)</td>
</tr>
</tbody>
</table>

* 95% confidence intervals
Table 5.2.: AIC for the linear mixed models to test which characteristic explains the change in breeding performance best

<table>
<thead>
<tr>
<th>Model</th>
<th>Females</th>
<th>Males</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1669</td>
<td>1750</td>
</tr>
<tr>
<td>Intercept and linear slope</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>1671</td>
<td>1752</td>
</tr>
<tr>
<td>(Age)$^2$</td>
<td>1671</td>
<td>1751</td>
</tr>
<tr>
<td>Age: second-order polynomial</td>
<td>1673</td>
<td>1752</td>
</tr>
<tr>
<td>Number of past breeding attempts</td>
<td>1670</td>
<td>1749</td>
</tr>
<tr>
<td>(Number of past breeding attempts)$^2$</td>
<td>1671</td>
<td>1750</td>
</tr>
<tr>
<td>Number of past breeding attempts: second-order polynomial</td>
<td>1672</td>
<td>1750</td>
</tr>
<tr>
<td>Number of past successful breeding attempts</td>
<td>1664</td>
<td>1736</td>
</tr>
<tr>
<td>(Number of past successful breeding attempts)$^2$</td>
<td>1670</td>
<td>1746</td>
</tr>
<tr>
<td>Number of past successful breeding attempts: second-order polynomial</td>
<td>1660</td>
<td>1728</td>
</tr>
<tr>
<td>Pair bond length in years</td>
<td>1671</td>
<td>1751</td>
</tr>
<tr>
<td>(Pair bond length in years)$^2$</td>
<td>1671</td>
<td>1751</td>
</tr>
<tr>
<td>Pair bond length in years: second-order polynomial</td>
<td>1673</td>
<td>1753</td>
</tr>
<tr>
<td>Factors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>1678</td>
<td>1749</td>
</tr>
<tr>
<td>Age: 2 groups (x≤14, x&gt;14)</td>
<td>1671</td>
<td>1751</td>
</tr>
<tr>
<td>Number of past breeding attempts</td>
<td>1677</td>
<td>1749</td>
</tr>
<tr>
<td>Number of past breeding attempts: 2 groups (x=0, x&gt;0)</td>
<td>1667</td>
<td>1744</td>
</tr>
<tr>
<td>Number of past successful breeding attempts</td>
<td>1656</td>
<td>1730</td>
</tr>
<tr>
<td>Number of past successful breeding attempts: 2 groups (x=0, x&gt;0)</td>
<td><strong>1643</strong></td>
<td><strong>1714</strong></td>
</tr>
<tr>
<td>Pair bond length in years</td>
<td>1664</td>
<td>1755</td>
</tr>
<tr>
<td>Pair bond length in years: 2 groups (x=1, x&gt;1)</td>
<td>1671</td>
<td>1750</td>
</tr>
</tbody>
</table>

The models are fitted to the change in relative reproduction. They all have random intercepts for the individuals and either age, the number of past breeding attempts, the number of past successful breeding events, or the pair bond length as fixed effect. The models are fitted via maximum likelihood (ML).
CHAPTER 6

Aggregate Versus Average Individual Change of Cognitive and Physical Functioning at Old Age

This chapter will be submitted for publication under the authorship of Maren Rebke and Zhen Zhang.

Co-author contributions: Zhen Zhang had the idea to analyse data from the Chinese Longitudinal Health and Longevity Survey, he got hold of the dataset and organised it. He further commented on the manuscript.
6.1. Summary

In the light of an increasing life expectancy (Oeppen and Vaupel 2002) it is of growing importance to study the trajectories of cognitive and physical functioning in old age to assess potentially increasing health care costs. As heterogeneity within a population can produce dynamics at the population level that are different from the dynamics at the individual level (Vaupel and Yashin 1985) it is important to differentiate between conclusions based on the aggregate population level and average individual change. Christensen et al. (2008) provide a clear example of differing conclusions for individuals and the society depending on whether more homogeneous subgroups or the whole population is considered. We apply in this study a novel decomposition method described in Rebke et al. (2010) to two different datasets dealing with functioning at old age. We first reanalyse the results on the ability to function independently in daily living at old age for the Danish 1905-cohort given by Christensen et al. (2008). Secondly, we apply it to data on the Mini-Mental State Examination (MMSE) from the Chinese Longitudinal Health and Longevity Survey (CLHLS) to assess the change in the cognitive functioning over age of oldest old people (age 80 and above). The results lead to different conclusions for the individual and society: Both studies reveal a decline in average individual performance at old age. In contrast, the decline at the population level is less strong due to the tendency of individuals with lower cognitive and physical ability to leave the study earlier.

6.2. Introduction

Best-practice life expectancy has increased linearly since 1840 and there is no sign of slow down in this trend. After 1950 improvement in survival at older ages (65 years and older) contributed most to the increasing life expectancy (Oeppen and Vaupel 2002), causing a fear of potentially increasing health care costs associated with the longer lives of old people. The question arose whether longer life expectancy is connected to a longer time period of healthy life (compression of morbidity, Fries 1980) or disabled life time (expansion of morbidity, Olshansky et al. 1991). If it is assumed that a decrease in mortality will save the lives of only some mostly unhealthy people, who would have otherwise died, longevity conclusively leads to a higher level of disability. If on the other hand people live not only longer but also healthier, a higher longevity will not change the level of disability. Further improvements in medicine and understanding of human biology might even lead to
a reduction in disability in old ages (Pardes et al. 1999).

To add to the discussion of disabled life time at old age we are studying trajectories rather than looking at prevalence (the proportion of people being disabled) or incidence (how many are becoming disabled newly) at a fixed point in time based on survey data with several follow-ups. This deepens our understanding of the course of mental and physical functioning and ultimately helps to assess the health care costs.

The problem in the analysis of trajectories is that conclusions about individual characteristics can very often not be drawn directly from observations made at the aggregate population level due to a compositional change in the population (Vaupel and Yashin 1985). This compositional change can be caused by two different mechanisms: 1) by selective disappearance of individuals from the study population due to mortality, emigration or non-response or 2) by selective appearance of new individuals in the dataset due to an inclusion of new people in a follow-up of a survey because of immigration or to compensate the loss of studied individuals. Therefore it is important for this study to differentiate between conclusions based on the aggregate population level and average individual change.

Christensen et al. (2008) provide a clear example of differing conclusions for individuals and the society depending on whether more homogeneous subgroups or the whole population is considered. Their study of the complete Danish 1905-cohort revealed that the proportion of the population able to function independently declined by 6% from 39% to 33% during four adjacent assessments from 1998 to 2005. Only a minority of the people of this cohort were able to take care of themselves, but this minority was almost constant over time. However, when considering the more homogeneous subgroup of individuals who were still alive at the survey in 2005, at age 99/100, the proportion of independence declined from 70% at age 92/93 to 33% at age 99/100. This discrepancy was due to the higher mortality of the people depending on help (Christensen et al. 2008).

We apply in this study a novel decomposition method described in Rebke et al. (2010) to two different datasets dealing with functioning at old age. We first re-analyse the results on the ability to function independently in daily living at old age for the Danish 1905-cohort given by Christensen et al. (2008). Secondly, we apply it to data on the Mini-Mental State Examination (MMSE) from the Chinese Longitudinal Health and Longevity Survey (CLHLS) to assess the change in the cognitive functioning over age of oldest old people (age 80 and above). In longevity studies, the MMSE index is frequently used to assess the cognitive mental status in older adults. For this application the decomposition approach allows us to answer
whether the mental ability in the population stays constant, declines or improves over old age and if the observed pattern at the aggregate level is really a change in the individuals or just an artefact due to compositional change. Both datasets are particularly valuable because they include a large number of individuals. Most other studies analysing mental or physical functioning either focus on earlier ages or have a much smaller dataset.

### 6.3. Methods

In this study we are applying a straightforward approach to decompose and quantify the observed change at the aggregate population level into average individual change and compositional change that was developed by Rebke et al. (2010). The approach uses longitudinal data. It extends previous decompositions of change (Price 1970; Vaupel and Canudas-Romo 2002; Coulson and Tuljapurkar 2008; Kerr and Godfrey-Smith 2009).

#### 6.3.1. Equations

Rebke et al. (2010) show that population change $P$ can be exactly decomposed into average change for the surviving individuals $s$ (in terms of providing a measure at the next measurement occasion), which we call in this study average individual change, plus compositional change due to selective disappearance $d$:

$$ P = s + d, \quad (6.1) $$

where $P = V_{x+k} - V_x$ is the difference between the averages of a characteristic of all individuals at one instance of an explanatory variable (e.g. age; $V_x$) and the next instance ($V_{x+k}$) with $k$ being the considered interval, $s = v_{x+k} - v_x$ gives the respective difference for the survivors at one instance ($v_x$) and the next instance ($v_{x+k}$) and $d = v_x - V_x$ is the difference between the two means at the first instance. The change due to selective disappearance is zero if the population is homogenous or if no individuals drop out. Equation 6.1 can be extended to account for new individuals entering the study at the second instance. In this case a term representing a compositional change due to selective appearance of new individuals $a = V_{x+k} - v_{x+k}$ has to be added to equation 6.1:

$$ P = s + d + a. \quad (6.2) $$
This approach works for means in general. In the special case of a binary variable, as for example independence, the mean number of successes, or in other words the proportion of successes can be used for $V_x$, $V_{x+k}$, $v_x$ and $v_{x+k}$ (Rebke et al. 2010).

### 6.3.2. Data

The data of the Danish 1905-cohort were gathered in 4 waves in a longitudinal multi-assessment survey from 1998 to 2005 and include 2234 individuals. People born in 1905, being still alive in 1998 and living in Denmark were included. Whether a person is independent or dependent was determined by physical and cognitive functioning and symptoms of depression. Independence was defined as being able to perform activities of daily living without the help of others and having a MMSE score of at least 23 out of a maximum of 30, which indicates not being cognitively impaired. The details of the study are given in Christensen et al. (2008). We decompose the population change for the cohort from one wave of the survey to the next into the average change of the surviving individuals and the change due to a compositional change.

The CLHLS data were drawn from the Study No. 24901 of the Inter-University Consortium for Political and Social Research (www.icpsr.umich.edu; Zeng et al. 2002). The study was carried out on subjects aged 80 and above in 1998 and in three subsequent follow-up waves during 2000, 2002 and 2005. It was conducted in randomly selected counties and cities of 22 Chinese provinces, which represent 85% of the population (Zeng and Vaupel 2004). For each centenarian that volunteered to take part in the study one octogenarian and one nonagenarian in the surrounding area (ranging from the same village to the same county depending on availability) was interviewed. The oldest old were thus oversampled by not following a proportional sampling design. The MMSE index was computed from the total score obtained by an individual on the Chinese version of the MMSE questionnaire. The items of this questionnaire were tests assessing orientation, attention, immediate and short-term recall, language, and the ability to follow simple commands. The Chinese MMSE includes 23-items which are an adjusted version of the standard 30-items MMSE (Folstein et al. 1975) to make the questions more understandable and answerable for ordinary oldest old Chinese, the majority of whom are illiterate (Zeng and Vaupel 2002). During the interview, each item received a binary score, namely 1 for a correct answer and 0 if the answer was not correct. We use the questionnaire total score for our analysis. More details about the study are given in Zeng et al. (2002).

In this example we decompose the aggregate change into individual change, com-
positional change due to dropping out of individuals either by dying or loss for the follow-up (for example caused by migration or refusing to participate) and change due to new appearance. This new appearance of individuals in the study was due to new individuals agreeing to participate in the survey for the first time at a later wave than the first in 1998. We only include individuals that do not have a missing overall MMSE score in the first year of their participation in the survey. The data include information on survival and the results of the MMSE for 18,255 individuals. Whenever an individual has a missing MMSE score in one of the follow-up waves it is regarded as lost for the survey from that wave onwards. We first group the data in broad age classes: octogenarians (age 80-89), nonagenarians (age 90-99) and centenarians (age 100+) in the year 1998 or 2, 4, or 7 years older in the respective follow-ups. Further, the people at age 100 in the wave 1998 or 102, 104 or 107 in the respective follow-up waves were followed in detail. In this example we use jackknifing to estimate 95% confidence intervals (CI).

6.4. Results

Our results for the Danish 1905-cohort show that the changes in the proportion of independent individuals from one assessment to the next at the aggregate population level are small (−3%, −4% and 1% respectively; figure 6.1), starting from an initial level of 39% independent individuals at the ages 92/93. However, the proportion of independent individuals among the survivors decreases strongly over age. From ages 92/93 to 94/95 the proportion of independent individuals among the survivors decreases by −17%, from ages 94/95 to 97/98 by −17% and from ages 97/98 to 99/100 by another −15%. This tendency of a decreasing independence among the surviving individuals is not obvious at the aggregate population level because of a compositional change over age. Individuals who are dependent on others tend to die earlier than independent individuals. This selective mortality explains a change in the proportion of independent individuals of 14%, 13% and 16% from one considered age to the next.

For the Chinese data there is a decline in mental ability measured by the MMSE also consistently visible at the population level. For the age group 80-89 the MMSE declines from the initial score of 19.70 by −0.99 (CI: ±0.17), −1.10 (CI: ±0.18) and −0.45 (CI: ±0.28) respectively from one wave to the next (figure 6.2). The change at the population level from one wave to the following for the age group 90-99 is −0.87 (CI: ±0.28), −1.62 (CI: ±0.28) and −0.40 (CI: ±0.47; figure 6.3), with an initial score of 16.56 and for the age group 100+ the change from the initial score
Figure 6.1.: Change in percentages of the Danish 1905-cohort classified as independent between four assessments from 1998 to 2005 which correspond to the ages 92/93 and 99/100 for the considered cohort. Males and females are combined. The dark blue bars show the change at the aggregate population level, the light blue bars the change in independence among the surviving individuals, and the orange bars represent the change that is explained by selective disappearance.

The average individual decline is even stronger: for the age group 80-89 it is \(-1.31\) (CI: \(\pm 0.19\)), \(-1.68\) (CI: \(\pm 0.19\)) and \(-1.46\) (CI: \(\pm 0.28\); figure 6.2) for the age group 90-99 \(-1.77\) (CI: \(\pm 0.35\)), \(-2.35\) (CI: \(\pm 0.36\)) and \(-2.49\) (CI: \(\pm 0.48\); figure 6.3) and for the age group 100+ \(-2.25\) (CI: \(\pm 0.58\)), \(-1.53\) (CI: \(\pm 0.81\)) and \(-2.53\) (CI: \(\pm 2.07\); figure 6.4) respectively for the changes between the four different assessments. This discrepancy between the population and the average individual change is due to a strong selection effect with people having a low mental ability tending to leave the study earlier either by death or loss for the survey. This leads to a change in the composition of the population with an increasing proportion of individuals with a high MMSE score. The compositional change due to selective disappearance comprises \(0.41\) (CI: \(\pm 0.10\)), \(0.68\) (CI: \(\pm 0.11\)) and \(1.01\) (CI: \(\pm 0.17\)) for the age group 80-89 (figure 6.2) for the respective changes between the different waves of the survey,
1.30 (CI: ±0.23), 1.45 (CI: ±0.24) and 2.09 (CI: ±0.34) for the age group 90-99 (figure 6.3) and 2.31 (CI: ±0.41), 1.29 (CI: ±0.61) and 1.12 (CI: ±1.43) for the age group 100+ (figure 6.4). The difference between the average individual and the population change is especially evident in the oldest age group 100+, where the change at the population level is very small. Only the population change between the waves in 2002 and 2005 seems to be bigger but the confidence intervals are broad due to a small sample size.

The change due to new appearance of people in the study plays in all three age groups only a small role: −0.08 (CI: ±0.14) and −0.10, (CI: ±0.10) for the age group 80-89 (figure 6.2) for the change between the waves in 1998 and 2000 and between 2000 and 2002, −0.41 (CI: ±0.27) and −0.72 (CI: ±0.25) for the age group 90-99 (figure 6.3) and −0.07 (CI: ±0.27) and −0.08 (CI: ±0.40) for the age group 100+ (figure 6.4). There were no new individuals recruited for the last follow-up of the survey in 2005.

When going from the broad age groups to the specific age of 100 the picture stays similar (figure 6.5). The average individual mental ability declines strongly over age (−2.41 (CI: ±0.85), −1.29 (CI: ±1.25) and −2.68 (CI: ±2.54) from one wave to the next). But there is also a strong compositional change due to selective disappearance (2.70 (CI: ±0.59), 1.20 (CI: ±0.91) and 2.06 (CI: ±1.86)) leading to only a small change at the population level (−0.11 (CI: ±0.73), −0.50 (CI: ±0.98) and −0.62 (CI: ±2.40)) starting from an initial score of 13.72. The population change is for the age 100 particularly small. The change due to new individuals entering the study in the waves 2000 and 2002 is again not important (−0.40 (CI: ±0.46) and −0.42 (CI: ±0.67)). The confidence intervals when considering the specific age of 100 are however broad due to a relatively small sample size.

6.5. Discussion

Our approach decomposes the aggregate change into its components. This leads to two different conclusions for the Danish dataset on the independence of old people: one at the societal level and one at the individual level. It shows that the proportion of independent individuals indeed declines with age but that at the same time the composition of the population changes, leading to only a small change at the aggregate level. These results are consistent with the findings by Christensen et al. (2008).

Our results for the Chinese dataset point in the same direction. Individuals decline strongly in their cognitive ability with age measured by the MMSE score. Although
the population change is not as negligible as it is for independence in the Danish dataset there is again a much stronger decline at the individual level. However, this decline is counterbalanced by the tendency of individuals with lower mental ability leaving the study earlier. In summary, the prospect for the society is much brighter than for each individual, with a strong decline in functioning at old age for the individual, but a smaller decline at the population level.

In terms of health care costs this is therefore a positive prospect for the society but a bad outlook for the individual. Previous findings of a decline in the prevalence of disability in the last decades add to the positive prospects for the society (Freedman et al. 2002, 2004; Manton et al. 2006). Medical advances can be expected to reduce disability connected to diseases as they have done so in the past (Cutler 2003).

Interestingly, the results for both datasets go in the same direction although we are studying different aspects of a change in disability with age for very different
Change of Cognitive and Physical Functioning at Old Age

Figure 6.3.: Change in MMSE score for the age group 90-99 at the population level (dark blue bars) and the contributions to this change due to average individual change (light blue bars) and compositional change due to selective disappearance (orange bars) and new appearance of individuals in the study (yellow bars). The number of individuals considered in the three change intervals is 4888, 5501 and 3582 respectively. The error bars represent 95% confidence intervals.

countries. On the one hand, the first dataset comes from a developed country, as the information from most previous studies. On the other hand, China is a country with an emerging economy.

Our study extends the knowledge from previous studies on change in functioning for the oldest old. Only a few studies so far have targeted the oldest old (Hassing et al. 2002b; Kliegel and Sliwinski 2004; Kliegel et al. 2004; Oksuzyan et al. 2010). Most previous analyses deal with younger age groups. Given that the oldest old are the fastest growing group of the population in developed countries and the most likely of needing medical care (Zeng et al. 2002; Christensen et al. 2009; Corrada et al. 2010), they are of importance for health care planning. The strength of the Danish 1905-cohort and the CLHLS datasets is the large sample size for the oldest old. There is usually only limited information for oldest old because of mortality as well as loss of follow-up and consequently small datasets. While other datasets
Figure 6.4.: Change in MMSE score for the age group +100 at the population level (dark blue bars) and the contributions to this change due to average individual change (light blue bars) and compositional change due to selective disappearance (orange bars) and new appearance of individuals in the study (yellow bars). The number of individuals considered in the three change intervals is 2115, 931 and 385 respectively. The error bars represent 95% confidence intervals.

Many previous longitudinal studies provide evidence of a decline in cognitive functioning with age. Data from people aged 2-95 in the U.S. showed that cognitive functioning first rises at early ages and then declines after around age 20 (McArdle et al. 2002). An increasing age predicts a decline in performance on memory tasks in a longitudinal study over 16 years in people aged 55-81 at baseline in California (Long Beach Longitudinal Study). But age is not the only predictor of memory change in this study – there are other factors leading to individual differences like gender and changes in reasoning or vocabulary (Zelinski and Stewart 1998). The very select group of Catholic nuns, priests and brothers from the longitudinal Religious Order
### Change of Cognitive and Physical Functioning at Old Age

#### Study in the U.S.

The study in the U.S. also reported a declining pattern of cognitive functioning in individuals without dementia or Alzheimer’s Disease when at least age 65 old (Barnes et al. 2003; Wilson et al. 2003). In the Netherlands the Longitudinal ageing Study Amsterdam showed that fluid intelligence declines with age for people of age 60 and above (Aartsen et al. 2004) and in the Dutch Longitudinal Study Among the Elderly there was a decline in cognitive functioning in people aged 65-84 at baseline (Deeg et al. 1990). In the Swedish Adoption/Twin Study of ageing for ages 50 to 96 different measures of cognitive ability declined after age 65 (Finkel and Pedersen 2004) in agreement with the OCTO (Origins of Variance in the Old-Old) twin study where age predicted a decline in cognitive functioning for Swedes at least 80 years old (Johansson et al. 2004). Hassing et al. (2002b) found a decline in face recognition and object recall over the 3-year study period for people from the Kungsholmen Project in Sweden initially aged 90-101. For people older than 65 from Narón Council in

#### Table 6.5.

<table>
<thead>
<tr>
<th>Year of Survey</th>
<th>Change in MMSE Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>2002-2005</td>
<td>2002-2005</td>
</tr>
</tbody>
</table>

The error bars represent 95% confidence intervals.

#### Figure 6.5.

Change in MMSE score for the age group 100 at the population level (dark blue bars) and the contributions to this change due to average individual change (light blue bars) and compositional change due to selective disappearance (orange bars) and new appearance of individuals in the study (yellow bars). The number of individuals considered in the three change intervals is 903, 463 and 203 respectively. The error bars represent 95% confidence intervals.

---

**References:**

- Barnes et al. 2003
- Wilson et al. 2003
- Aartsen et al. 2004
- Deeg et al. 1990
- Finkel and Pedersen 2004
- Johansson et al. 2004
- Hassing et al. 2002b
Spain the MMSE score correlated negatively with age (Millán-Calenti et al. 2009).

There are only a few studies reporting no or only slight decline in cognitive functioning with age. The French longitudinal Paquid study of at least 65 year old people without dementia who lived at home at the time of the baseline survey showed only a slight decline in MMSE-score (Jacqmin-Gadda et al. 1997). Kliegel and Sliwinski (2004) found several different trajectories of cognitive functioning in the Heidelberg Centenarian Study in Germany but an overall analysis revealed for these 100 year old individuals that mental ability was stable (Kliegel et al. 2004).

While many previous studies excluded people with dementia, apart from some exceptions (Deeg et al. 1990; McArdle et al. 2002; Kliegel and Sliwinski 2004; Kliegel et al. 2004; Millán-Calenti et al. 2009), we include individuals with dementia in our analysis. A study of Americans aged 90 and older showed that incidence rate of dementia increased exponentially with age from 12.7% in the ages 90-94 to 40.7% for the ages 100+ (Corrada et al. 2010). We therefore think that dementia is an important factor for the change in cognitive functioning and part of the ageing process.

Apart from cognitive ability, physical functioning is part of the measure of independence. In accordance with the average decline in independence for survivors there is evidence from previous studies for a decline in physical functioning with age. In the Danish 1905-cohort strength decreased linearly over age, with individuals having a higher initial strength decelerating faster (Oksuzyan et al. 2010). In the Swedish OCTO twin study of individuals aged 79-96 age explains performance in several physical tasks at baseline. There was a linear decline in grip strength and chair-stand performance. The decline in peak expiratory flow even accelerated over the course of this longitudinal study (Proctor et al. 2006).

The Established Populations for Epidemiologic Studies of the Elderly in the U.S. showed that higher ages predict more disability at baseline (Taylor and Lynch 2004). Further, there was a decline with age in physical functioning for people aged 65 and older. The rate of decline increased with advancing age at baseline (Beckett et al. 1996; Taylor and Lynch 2004). The selected group of employed males and unmarried employed females from the Longitudinal Retirement History Study in the U.S. of people aged 58-63 in 1969 also showed that the age trajectory of functional impairment is non-linear (Clark and Maddox 1992). However, Maddox and Clark (1992) stated that there are differences in trajectories of impairment for this dataset which cannot simply be explained as a function of age. Other factors were important too, like sex, education and economic status. Although there was a general association of disability and age for people between age 15 and 74 living in Eindhoven and
surroundings in the study, Nusselder et al. (2006) also found differences between individuals in their trajectories of disability – the authors identified 9 different trajectories. Age further predicts disability in the Rotterdam Study of people aged 55 and above (Taş et al. 2007).

In terms of prevalence of disability in our datasets, previous analyses showed that in the Danish 1905-cohort 46% were not impaired, 32% were mildly impaired and 22% were severely impaired in 1998 (Andersen et al. 2002). In a cross-sectional analysis of the CLHLS survey in 1998 the percentage of people in good cognitive functioning measured by MMSE dropped from 79.83% in the age group 80-84 to 25.53% in the age group 100-105. The proportion of people with no disability in terms of activities of daily living declined from 83.55% in the age group 80-89 to 36.77% in the age group 100-105 (Zeng and Vaupel 2002).

Our results show for both datasets that there is a strong positive compositional change due to selective disappearance, meaning that individuals not present at the next follow-up are on average worse in their cognitive performance and are more frequently dependent than the remaining individuals. A relationship between cognitive or physical functioning and longevity was also shown in previous studies. We were only studying the relationship of leaving the study population and the level of performance. In the literature there are contradicting results whether a low level of functioning or an accelerated decline in function predicts subsequent survival.

For cognitive functioning some studies showed an association of mortality and the level of cognitive ability as well as the slope of the loss of functioning: an increased rate of decline in cognitive functioning was evident about 43 months prior to death in Catholic nuns, priests and brothers from the Religious Orders Study in the U.S. Further, the average cognitive functioning at baseline was lower for individuals dying during the study than for survivors (Wilson et al. 2003). In the Seattle Longitudinal Study of ages between 22 and 95 people who subsequently died performed worse at their last measurement in several cognitive tasks compared to survivors for the age group above 75 but not for younger age groups. In addition, the decline in verbal meaning and psychomotor speed between the last two measurements was stronger for descendants than for survivors (Bosworth and Schaie 1999).

Other studies support only an association of mortality and the level of cognitive functioning. A study based on the Danish 1905-cohort also used in our study provides support for this association. Andersen et al. (2002) found that mortality risk within two years increased with cognitive impairment measured by MMSE and a cognitive composite of five tests.

The Swiss Interdisciplinary Longitudinal Study on the Oldest Old showed that
verbal fluency performance level of individuals aged 79 to 85 predicted survival, whereas cognitive change was not predictive (Ghisletta 2008). In the Swedish Kungsholmen Project MMSE score at baseline for people aged 75 and older was lower for people being close to death compared to survivors (Laukka et al. 2006). For individuals initially aged 90-101 the object recall performance at baseline of survivors was better than for the individuals that were dying subsequently, but there was no effect of the performance in face recognition, word recall and word recognition. Although there was a decline in face recognition and object recall found in this study over a 3-year period, the slope was not dependent on subsequent mortality (Hassing et al. 2002b). In accordance, the OCTO twin study showed for octogenarians an association between the baseline cognitive performance and subsequent survival (Johansson and Zarit 1997; Hassing et al. 2002a) but no accelerated decline before death (Hassing et al. 2002a). Korten et al. (1999) found that poor cognitive functioning was a predictor of mortality in people aged 70 and older from Australia. The U.S. Atherosclerosis Risk in Communities Study of adults aged 48-67 showed that cognitive functioning at baseline predicts further life expectancy (Pavlik et al. 2003). Cognitive impairment was related to morbidity and mortality for people older than 65 in Spain (Millán-Calenti et al. 2009). In contrast to this evidence for a survival effect of the level of mental functioning, there was no significant association of cognitive ability at baseline and subsequent survival for people of age 100 in the Heidelberg Centenarian Study (Kliegel et al. 2004).

There are further studies supporting only the association of the slope of decline in cognitive functioning and mortality: for Dutch people age 70 and above, a strong decline in cognitive functioning was associated with a short survival time (Deeg et al. 1990). Data of people aged 65 and over from the Monongahela Valley Independent Elders Survey in the U.S. also showed that cognitive decline predicted subsequent survival (Lavery et al. 2009). In the UK Health and Lifestyle Survey a decline in cognitive performance over the 7-year study period was associated with an increased risk of dying for the age group 60+ (Shipley et al. 2007). Schupf et al. (2005) studied both cognitive and physical functioning in people at least age 65 from Manhattan and found that those with a strong decline in different measures of functioning had a higher risk of death than people with slower or no decline. The slope effect for mortality was stronger in people age 75 and below than for older participants. However, Laukka et al. (2006) showed that in their study of individuals aged 75 and above from the Kungsholmen Project the accelerated decline in functioning prior to death vanished when individuals with dementia were excluded, suggesting that it is due to the impact of preclinical dementia.
For physical functioning there is additional evidence for an association with mortality of both the level and the slope of decline in function, including results from the datasets used in our study: Oksuzyan et al. (2010) showed previously that longer life was associated with higher initial strength and a slower deceleration in strength in the Danish 1905-cohort alive in 1998. Shen and Zeng (2010) found in the 2002-2005 sample of the CLHLS of individuals at least aged 65 with a large subsample of people aged 85 and older that a greater resilience is associated with a lower mortality risk. This effect was stronger for ages below 85. A high resilience was probably related to better health leading to a lower mortality risk (Shen and Zeng 2010).

Performance in activities of daily living declined in the final year before death for participants aged 65 and older in the Established Populations for Epidemiologic Studies of the Elderly in the U.S. (Lunney et al. 2003). Further, individuals who did not survive the 5-year study period declined faster, started from a lower level of cognitive functioning at baseline and were more likely to become disabled (Beckett et al. 1996). For Jewish long-term care residents aged 65 and above in Massachusetts decline in activities of daily living was greatest 3 month prior to death but depended on the underlying disease (Chen et al. 2007). Physical disability is connected to a higher death rate for individuals of at least 55 in California (Oman et al. 1999). For Australians aged 70 and older poor physical health was a predictor of mortality (Korten et al. 1999).

A strength of our study is that both datasets are population based. Many other studies analysing the change in cognitive or physical functioning with age excluded institutionalised individuals (Beckett et al. 1996; Jacqmin-Gadda et al. 1997; Zelinski and Stewart 1998; Johansson et al. 2004; Nusselder et al. 2006; Proctor et al. 2006; Millán-Calenti et al. 2009). However, this part of the population should be included in order to avoid biasing the results. At least of the Danish 1905-cohort more than a third are living in institutions (Christensen et al. 2008). Further, 53% of elderly (65 and above) being cognitive impaired or limited in daily activities are living in institutions in England and Wales (Melzer et al. 1999).

In cross-sectional studies it is not possible to distinguish between age effects and cohort effects. Therefore it is important to study longitudinal datasets. A longitudinal approach to study trajectories over age now seems to be common, given the number of studies discussed above. Most authors of these studies tried to account for repeated measures of individuals in their models using a random effects model or growth curve model (Clark and Maddox 1992; Maddox and Clark 1992; Beckett et al. 1996; Jacqmin-Gadda et al. 1997; Hassing et al. 2002b; McArdle et al. 2002; Barnes et al. 2003; Wilson et al. 2003; Aartsen et al. 2004; Finkel and Pedersen 2004;
Johansson et al. 2004; Taylor and Lynch 2004; Proctor et al. 2006; Oksuzyan et al. 2010). While some studies acknowledge that it is important to consider a potential selection bias if there is attrition on the variable of interest (e.g. Deeg et al. 1990; Maddox and Clark 1992; Beckett et al. 1996; Bosworth and Schaie 1999; Johansson et al. 2004; Kliegel et al. 2004; Oksuzyan et al. 2010), none of them account for it in the estimation of change over age. However, there are a few studies that account for attrition in the analysis of change in functioning with time (Ten Have et al. 2000; Jacqmin-Gadda et al. 2010). The same methods could be applied also to changes over age.

Previous studies were not able to quantify the effect of compositional change. This is the main advantage of the approach that we used in this study. In addition, as shown in the analysis of the CLHLS we were able to include new respondents at follow-up waves and estimate their effect of selective appearance for the population-level change. In the discussed previous longitudinal studies of change in functioning with age new respondents were not included.

There are some limitations to our study. For the Danish 1905-cohort we only have the summary data published by Christensen et al. (2008). Therefore it was not possible to calculate the confidence intervals for our results. Further, we could not distinguish in our analysis between causes of loss of follow-up, i.e. dead, emigrated or refusing to participate. But Christensen et al. (2008) showed for the data on the Danish 1905-cohort used in this study for the analysis of change in independence that correcting for missing data due to nonparticipation does not change the results. They came to the conclusion that the difference between the change at the population level and the individual level is mainly caused by selective mortality.

We showed that our approach is useful for survey studies where it is important to distinguish between the actual changes of the respondents and the effects caused by the dropping out or entering of certain respondents. We were able to assess the change in the level of mental ability and proportion of independent people over age on the aggregate level and to decompose this change into average individual change and the contribution of compositional change. We showed that this leads to two different conclusions: one on the societal level and one on the individual level.
CHAPTER 7

Discussion
7.1. Overall Discussion

7.1.1. Overall Findings

In this thesis age-specific patterns of performance are analysed to gain insight into the mechanisms of ageing. What is found continuously throughout this thesis is an increase in reproductive performance in early ages and a tendency for a decline in old ages for the common tern (see chapter 2, 3 and 5). Although reproductive senescence at advanced ages is not found under all circumstances (see chapter 2), it is the predominant finding.

The hump-shaped pattern of age-specific reproductive performance is widely believed to be the usual shape in wild populations of bird and mammal species (summaries in Clutton-Brock 1988; Newton 1989; Wooller et al. 1992; Forslund and Pärt 1995). An increase in early life and senescence in older life further follows the predictions from life history theory models estimating the optimal strategies by optimal allocation of resources (Pianka and Parker 1975; Cichoń 2001; Kindlmann et al. 2001). The improvement in reproductive performance clearly contradicts the prediction from the classic evolutionary theory of ageing in which performance is destined to decline from maturity onwards due to a decreasing selection pressure with age (Medawar 1952; Williams 1957; Hamilton 1966; Kirkwood 1977). The results of this thesis add to the view that real life history strategies are more complicated than the classic evolutionary theory of ageing suggests and that the field should move to models that are able to predict various performance patterns (Baudisch 2008) and not only senescence.

Although some differences in age-specific reproductive patterns between male and female common terns are found at advanced ages in chapter 2 and 5 in separate analyses of the sexes, the patterns are very similar when modelling all the processes leading to reproductive output and both sexes simultaneously (chapter 3). The absence of strong sex differences in reproductive patterns seems to be the typical result for monogamous vertebrates (Rattiste 2004; Wintrebert et al. 2005; Clutton-Brock and Isvaran 2007; Lecomte et al. 2010; Limmer and Becker 2010; Kim et al. 2011; Knape et al. 2011). In the hierarchical Bayesian model in chapter 3 unmarked partners are included and the sex is determined when unknown. In contrast, the analysis in chapter 2 and 5 only considers marked birds of known sex. The analysis in chapter 3 thus incorporates much more information. The difference in age-specific reproductive performance at old age might be an artefact due to the lower sample size of males in the older ages and a lower maximum age present in the population.
Chapter 2 and 3 show that there are persistent differences between individuals in reproductive performance, but the effect of selective disappearance of individuals, and even more so of selective appearance, is not strong compared to the average change within individuals (chapter 5). Therefore, at least for the common tern previous population-level analyses are broadly correct. Previous analyses could only speculate about the importance of the different factors to the population-level change and were not able to quantify these contributions. In contrast, there is a strong selective disappearance of individuals in the study in chapter 6 of the ability to live independently and of cognitive functioning at old age in humans masking a decline in functioning with age within individuals.

Variation in environmental conditions can have an effect on phenotypic traits (Altwegg et al. 2007; Ozgul et al. 2009). The analysis in chapter 2 shows that breeding conditions measured by the average reproductive success of the population have an effect on reproductive performance in the common tern. The effect of the quality of the year is found to be age-specific, with disproportionately high benefits of a good breeding year for intermediate ages. But overall the early increase with age in reproductive performance in the common tern is still much stronger than the difference between years of different quality.

The common tern is a monogamous species and consequently two partners breed together. There is a positive relation of adult life expectancy and mate fidelity (Briëd et al. 2003) and thus pair bond length might be of importance especially in long-lived species like the common tern. The effect of the pair bond length as an increasing function has never been studied in the common tern before, but previous studies of divorce did not give a concise picture (González-Solís et al. 1999a; Ludwigs and Becker 2007). The hierarchical Bayesian analysis in chapter 3 shows that only the age of both partners and their reproductive quality affects the reproductive output of the pair but not the length of time they have bred together in the past. Nevertheless, individual common terns are more likely to choose a partner that they know, regardless of the accumulated previous breeding outcome with that partner, suggesting that they make a choice once at the beginning of their breeding career and tend to stay with that partner.

In conclusion, the different analyses show that age is the important component driving the reproductive pattern of the common tern. However, for other traits and different populations this result might not be true.
7.1.2. Developed and Applied Methods

Appropriate methods handling the challenges complicating the accurate quantification of the age pattern of performance are developed and applied in this thesis.

In chapter 2 a proportional odds model with random effects (Tutz and Hennevogl 1996; Agresti 2002) is applied to the age-specific number of fledglings in the common tern. Although this is an existing statistical model appropriate for repeated ordinal response data, it has not been used previously in ecology and evolutionary biology. A comparison showed that especially when the process of not breeding is part of the analysis and the maximum number of offspring per breeding event has a low maximum, applying a Poisson model, which is based on more restrictive assumptions, is inappropriate and might lead to wrong results.

A hierarchical Bayesian model is developed in chapter 3 that includes all steps leading to the number of fledglings. These steps are survival, the decision to attend the breeding colony, the decision to breed, the choice of a partner, the production of a certain number of fledglings and the decision to breed repeatedly within a breeding season. This model sheds new light on various life history characteristics of the common tern. Several aspects are consistent with previous findings for this species, which validates this complex model.

While the proportional odds model with random effects in chapter 2 and the hierarchical Bayesian model in chapter 3 model differences between individuals they cannot quantify the contributions of selective effects to the population-level pattern. I develop in this thesis a mathematical decomposition of the change in a trait at the population level into average individual changes, changes due to selective disappearance of individuals and, with an extension, changes due to selective appearance. Chapter 4 shows how the decomposition without the extension can be derived from the well-known Price equation (Price 1970). The decomposition is described in detail in chapter 5. This straightforward method can be useful for a wide range of applications. In this thesis it is applied to reproductive performance of the common tern (chapter 5), mental and physical functioning at old age in two different human datasets (chapter 6) and various epidemiological as well as economic studies (appendix A), where the contributions of the different processes depend on the study. The decomposition method has further been used already in two published papers by other authors. Evans et al. (2011) decomposed the change with age in ornamental plumage traits and different measures of reproductive performance in the collared flycatcher (Ficedula albicollis) applying this method. Nussey et al. (2011) used the decomposition method to analyse the change in body mass with age in different ungulate species.
In conclusion, I present several useful methods to study changes in performance. Often ecologists and evolutionary biologists apply standard methods even though they might not account for potential bias or are inappropriate. But the results of this thesis highlight that choosing an appropriate model can be important. This might also be the case for other species and traits under study.

What has been presented here is a discussion covering all the topics of this thesis jointly. The results of the analyses within each chapter are discussed there.

7.2. Directions for Future Research

7.2.1. Extending the Hierarchical Bayesian Model

Although the hierarchical Bayesian model described in chapter 3 is already very complex and able to answer a range of questions there are several potential extensions.

First, I would like to test for different mortality functions. So far the widely used Gompertz function (Gompertz 1825) is implemented. Other functions that I would like to implement, and test whether the model fit is improved, are a Weibull (Weibull 1951), a Gompertz-Makeham (Makeham 1860), a Gamma-Gompertz (Vaupel et al. 1979), a Gamma-Gompertz-Makeham (Manton et al. 1986) and a Siler model (Siler 1979). The Weibull is another commonly used mortality function in demography. The Gompertz-Makeham has a term for age-independent mortality and thus shows if there is a strong effect of external mortality for the common tern. The Gamma-Gompertz model accounts for individual heterogeneity in mortality by combining the Gompertz mortality function with a Gamma function as a frailty distribution of the individuals. The Gamma-Gompertz-Makeham combines the properties of the Gompertz-Makeham and the Gamma-Gompertz model. The Siler model enables one to test whether mortality has a bath tub shape; with an initial decline at early ages and an increase at older ages. Further, I would like to test for differences between males and females in the mortality function. The model currently includes one mortality function for both sexes combined.

Second, instead of using the truncated Poisson function with an upper limit of 3 for the number of fledglings in the hierarchical Bayesian model in chapter 3 I would like to incorporate the proportional odds model with the individual random effects described in chapter 2. Although we separate the process of whether a bird breeds or not from the process of producing a certain number of fledglings in the hierarchical
model and truncate the Poisson distribution, the proportional odds model might still be more appropriate.

Third, I would like to link the survival function in the hierarchical model to the function for the number of fledglings. The introduction of this link will show if there is a trade-off between reproduction and survival or a selection effect and will account for that effect. The link could be implemented by either having the same individual effects in the survival and reproductive function (Wintrebert et al. 2005) or by correlating an individual effect in the reproductive function with an individual effect in the survival function (Cam et al. 2002; Knape et al. 2011).

7.2.2. Comparative Study

Applying the decomposition method derived in chapter 4 to age-specific reproductive performance of the common tern shows in chapter 5 that selective disappearance, and even more so selective appearance, are relatively unimportant compared to average changes within the individuals. I would like to apply the decomposition method to a range of different species of mammals and birds to test if this result is general and to compare the age-specific reproductive patterns. To apply the approach, long-term datasets with repeated information on a reproductive trait of individuals are needed. The recapture probability should be close to one. Potential species for which such excellent datasets exist are the following: bighorn sheep (*Ovis canadensis*), kittiwakes (*Rissa tridactyla*), mute swans (*Cygnus olor*), red deer (*Cervus elaphus*) and Soay sheep (*Ovis aries*).

7.2.3. Long-Term Effects of Partner Change

Even though there is no effect of pair bond length found on immediate breeding performance in chapter 3, the different partner change strategies might have a long-term effect. Changing the mate often or being faithful might influence the lifetime reproductive success or the age at death.

I would like to apply the concepts of dynamic heterogeneity and entropy introduced by Tuljapurkar et al. (2009), and extended by Steiner et al. (2010) accounting for survival, to the common tern dataset to get further insights into this topic. Caswell (2011) uses a similar approach applying Markov chains with rewards. There are two non-exclusive potential explanations for the observed variation in individual life histories, which can lead to the same observed pattern: differences between individuals fixed at birth (i.e. fixed heterogeneity) and dynamically generated stochastic variation of reproduction and survival (i.e. dynamic heterogeneity (Tuljapurkar et al. 2009)).
Discussion

2009) or individual stochasticity (Caswell 2011) as equivalent terms). While fixed heterogeneity describes unobserved persistent individual differences the concept of dynamic heterogeneity characterises differences in the life history trajectories of individuals which are generated by probabilistic movements between stages. Dynamic heterogeneity is thus caused by stochastic events (Tuljapurkar et al. 2009; Caswell 2011). The proportional odds model in chapter 2, the hierarchical Bayesian analysis in chapter 3 and the decomposition approach in chapter 5, show that there is some fixed heterogeneity between individuals. But there are also many different individual life histories and differences in survival among the individuals following the same life history. Chapter 2 shows that environmental conditions influence the reproductive performance and can thus cause variation in life histories between individuals.

The dynamic heterogeneity approach uses a stage-structured transition matrix with an absorbing Markov chain when incorporating survival into the model (Steiner et al. 2010; Caswell 2011). To study the effect of partner changes, the stages will be defined by reproductive performance in combination with the partner status. The transition matrix can be used to derive an estimate of the diversity in individual life histories, which is called entropy. Further, persistence within stages over the life of the individuals and the proportions of individuals in different stages can be calculated (Steiner et al. 2010). Mean, variance and skewness of lifetime reproductive successes can be calculated applying the concept of dynamic heterogeneity using Markov chains with rewards (Caswell 2011). The dynamic heterogeneity approach allows the analysis of how variation in reproductive and partner change trajectories affects the variance in fitness components like expected lifetime reproductive success and age at death. Variances within and between trajectories in lifetime reproductive success and age at death can be compared to get information about the importance of differential survival in the population (Steiner et al. 2010).

7.3. Conclusion

In this thesis I develop and apply new methods as well as applying existing methods to analyse age-specific performance and address challenges that complicate the accurate quantification of these patterns. The applications highlight that the use of appropriate methods might be of importance to get unbiased results. Several processes can influence the rate of age-specific changes and their contribution to the population-level pattern are specific to the trait and population studied. There is a strong influence of selective disappearance of individuals on functioning at old age in humans. Although factors like selective disappearance and the environment
influence the common tern, I find that age is the most important determinant of the individual reproductive pattern in this species. Overall, there is an increase in reproductive performance at early adulthood and a tendency for senescence at old age in the common tern.
Exploring Longitudinal Data Subject to Non-Monotone Missingness: 
a Simple Decomposition of Changes

This appendix will be submitted for publication under the authorship of James E. Oeppen and Maren Rebke.

Co-author contributions: James E. Oeppen chose the examples and performed the analysis after I taught him how to apply the decomposition method. James E. Oeppen wrote most of the results section, while I mainly wrote all the other parts.
A.1. Summary

This study summarises a recently developed approach by Rebke et al. (2010) to decompose population-level change into three contributing components: the average change within individuals, the compositional change due to selective dropout and due to selective recruitment. We illustrate the wide range of applications for this decomposition by applying it to four, publicly available, epidemiological and economic datasets on the number of doctor visits before and after the German medical funding reform, on the effect of drug treatments on schizophrenia, on childhood obesity and on unemployment and absenteeism in a German company. Our results show that depending on the study selective effects often cannot be ignored: The decline in doctor visits during the funding reform is mainly an effect of a selective dropout of individuals who went to the physician at an above average frequency. Selective dropout of individuals with severe schizophrenia symptoms masks a deterioration of the remaining study participants. The high absentee rate of people subsequently leaving the company offsets the declining rate for individuals staying employed. In the obesity study selective dropout causes a decline in obesity at the population level only at ages 11 to 14, while population and individual changes are coordinated at the other ages.

A.2. Introduction

Monotone missingness refers to the situation in longitudinal data where all subjects are observed at baseline, but some drop out and never return to observation. This process can be referred to as right-censoring, attrition, dropout, or loss to follow-up. In other cases it could also simply be death. Non-monotone missingness occurs when some subjects drop out and others enter, having missed one or more of the previous observation times. These may be subjects that experienced recruitment during the course of the study, termed left-truncation or panel refreshment, or subjects that were observed intermittently. In some cases the missingness pattern may be under experimental control, but generally this is not the case.

A crucial issue in the analysis of incomplete longitudinal data is understanding the processes that create missingness. These can be classified as: Missing Completely at Random (MCAR) – missingness does not depend on the observed or unobserved data. Missing at Random (MAR) – missingness does not depend on the unobserved data. Missing Not at Random (MNAR) – the missing process is ‘non-ignorable’ and the probability of being missing is dependent on the unobserved data, i.e. the
probability of missingness is related to the value of the responses that are missing. With MCAR data we can proceed with the analysis as if no data were missing. In analysing MAR data we are assuming that the missing observations are drawn from the same distributions as the comparable known data. In both cases the missing mechanism is ignorable, while in the case of MNAR it is not (Fitzmaurice et al. 2004). In practice it is often difficult to know which type of missingness holds.

If selective missingness within a heterogeneous population occurs, which is characterised as MNAR, the population averages are subject to a compositional change. Therefore, the pattern of changes within individuals and population-level changes can be very different (see Vaupel and Yashin 1985; for examples) and it is of importance to distinguish between both levels.

Rebke et al. (2010) developed a straightforward and simple approach to decompose the observed change at the aggregate population level exactly into average individual change and change due to dropout. It can be extended additionally by separating the change due to new subjects entering the study population from the other factors contributing to the population-level change.

The basic decomposition method can be related formally to the Price equation (Price 1970), which is shown in Rebke (2010). The Price equation was originally developed for population genetics and is a fundamental and well-known representation of changes in a trait value from one generation to the next (Price 1970; Gardner 2008). It is applicable to selection problems in general (Price 1970) and has now been applied in various other research fields (Andersen 2004; Day and Gandon 2006; Jäger 2008; Collins and Gardner 2009). Different extensions of the Price equation have been developed (Price 1972; Frank 1997; Grafen 2000, 2002, 2006; Heywood 2005; Coulson and Tuljapurkar 2008; Rice 2008; Kerr and Godfrey-Smith 2009). Kerr and Godfrey-Smith (2009; eqn. 1) presented an extension that can handle immigration into the studied population, which, although formulated differently, relates to the extended version of the decomposition of Rebke et al. (2010).

The advantage of the decomposition of Rebke et al. (2010) over the original Price equation and also the extension by Kerr and Godfrey-Smith (2009) to study average changes within individuals rather than changes between parents and offspring is the more intuitive formulation with clear meanings of the different parts of the equation. As survival and recruitment are binary Rebke (2010) showed that the covariance term in the Price equation, which gives the compositional change in a non-intuitive way, can be reduced to a difference term describing the change in a straightforward way. Like in Coulson and Tuljapurkar (2008) the parts of the equation are not all related to the original set of subjects at the first measurement occasion, it is instead
distinguished between a group of all individuals and a group of the subjects that do not drop out.

The decomposition uses longitudinal data with repeated measures and individuals dropping out or entering the study need to be distinguishable. The decomposition can be applied to continuous variables as well as to proportions that are recorded at fixed occasions. These occasions do not necessarily have to be regularly spaced, and would normally be time points, durations or ages, but they can also be ordinal stages, such as pregnancy order. It is also applicable to panel-count data and rates defined over a preceding interval. It is not suitable for irregularly observed data, unless they can be meaningfully converted to cross-sectional data or rates. It is an important feature of the method that it is suitable for data with non-monotone missingness.

To illustrate the decomposition method we have selected four studies in epidemiology and economics for which the datasets are publicly available. These are studies on the number of doctor visits before and after the German medical funding reform for women between 20-60 in full-time employment, on the effect of drug treatments on schizophrenia, on childhood obesity as well as on unemployment and absenteeism in a German company. Our analyses are not intended to be complete, but to illustrate how the decomposition might form part of an initial exploration of the data.

A.3. Methods

Change at the aggregated population level $P$ is exactly decomposable into average change within the subjects staying in the study $s$ and a compositional change due to selective disappearance of certain subjects $d$:

$$P = s + d.$$  \hspace{1cm} (A.1)

In this equation the population change is given by $P = V_{x+k} - V_x$. This is the difference between the average $V_x$ characteristic under study of all individuals at one instance $x$ and the average $V_{x+k}$ at the next instance $x+k$, where $k$ gives the interval between the two instances. Scaling the averages leads to proportions, which can consequently also be analysed with this equation. The average change for the subjects staying in the study is represented by $s = v_{x+k} - v_x$. This is the respective difference for these staying subjects at one instance ($v_x$) and the next instance ($v_{x+k}$). The term measuring change due to selective disappearance is given by $d = v_x - V_x$. This is the difference between the average for the staying subjects $v_x$ and the average for all subjects $V_x$ at the first instance $x$. The decomposition was
originally shown in Rebke et al. (2010), which also gives a graphical representation of equation A.1.

To account for new subjects entering the study population at a later occasion, it is possible to extend equation A.1 (Rebke et al. 2010). If a selected group of individuals with certain characteristics enters, this addition leads to another change in the composition of the population. This change is thus caused by selective appearance $a$ of subjects, which is captured by adding the term $a = V_{x+k} - v_{x+k}$ to equation A.1:

$$P = s + d + a.$$  \hfill (A.2)

Graphical illustrations of this extension are given in Rebke et al. (2010).

The theoretical distributions of the studied characteristics are unknown. Approximate 95% confidence intervals are calculated for the population change and the components leading to that change via bootstrapping with 999 iterations. The package `boot` (Davison and Hinkley 1997; Canty and Ripley 2009) from the statistical software R (R Development Core Team 2010) is used for the calculations.

### A.3.1. Data Sources

The data from example 1 concerning the number of doctor visits for German women between 20-60 in full-time employment is publicly available as the dataset `drvisits` from the R-package `sabreR`. The data from example 2 on the effect of drug treatments on schizophrenia can be downloaded from Prof. Peter Diggle’s website: [http://www.lancs.ac.uk/staff/diggle/APTS-data-sets/PANSS_short_data.txt](http://www.lancs.ac.uk/staff/diggle/APTS-data-sets/PANSS_short_data.txt). The dataset from example 3 on childhood obesity comes from the Muscatine Coronary Risk Factor Study (Woolson and Clarke 1984). It is given in Woolson and Clarke (1984; table 10) and can be downloaded from the website of the book by Fitzmaurice et al. (2004): [http://biosun1.harvard.edu/~fitzmaur/ala/obesity.txt](http://biosun1.harvard.edu/~fitzmaur/ala/obesity.txt). The data in example 4 on unemployment and absenteeism in a German company comes from a study by Kauermann and Ortlieb (2004) and can be downloaded from the Royal Statistical Society website: [http://www.blackwellpublishing.com/rss/Volumes/Cv53p2.htm](http://www.blackwellpublishing.com/rss/Volumes/Cv53p2.htm).
A.4. Examples

A.4.1. Example 1: The Impact of German Medical Funding Reform on Doctor Visits – a Two Panel Study

In 1997 Germany undertook a reform of public health-care provision that included raising the monetary costs to patients of obtaining prescribed medication and capping the public reimbursement budgets of physicians. To analyse the effect of these reforms on patient behaviour Winkelmann (2004) undertook a ‘before and after’ study using as an outcome variable the number of doctor visits per patient in the three months prior to interview in 1996 and 1998 – years chosen to straddle the mid-1997 reform. The data were obtained from the German Socio-Economic Panel (GSOEP) annual household survey (http://www.gsoep.de/en/soep) in which ‘doctor’ was defined to include both doctors and dentists. This longitudinal survey exhibits both attrition and recruitment, so the missingness pattern is non-monotone. The dataset used here follows the structure of the Winkelmann (2004) study but is restricted to women aged 20-60 in full-time employment - as analysed by Rabe-Hesketh and Skrondal (2008; chapter 5 ). Doctor visits data from the same survey have also been analysed by Riphahn et al. (2003) and Greene (2009), however, with a different research question in the latter. Greene (2009) was interested in the effect of public insurance and private add-on insurance on the number of doctor and hospital visits.

The observation patterns in the data are shown in table A.1, sorted by the number of cases having each pattern. The numbers of subjects in each panel are 1100 and 1127, but only 709 of the 1518 women are observed both before and after the reform, representing panel attrition and recruitment of about 36%. For all women aged 20-60 in West Germany at these dates, natural attrition from either mortality or reaching the 60th birthday amounted to about 2.5%. Natural recruitment through reaching age 20 between 1996 and 1998 was about 1.8%. Assuming these figures to be relevant to this subsample of working women, almost all the turnover between the two panels is likely to have been triggered by withdrawal for reasons other than normal retirement or mortality, with subsequent recruitment to maintain the panel size. If withdrawal is not independent of health, this level of turnover could have an important effect on changes in the average number of doctor visits. Two of the most likely causes of withdrawal from full-time female employment are unemployment and pregnancy, and both are expected to be associated with elevated medical usage before dropout.
Table A.1.: Doctor visits observation patterns

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>709</td>
<td>46.71</td>
<td>1</td>
<td>1</td>
<td>1418</td>
<td>63.67</td>
</tr>
<tr>
<td>2</td>
<td>418</td>
<td>27.54</td>
<td>.</td>
<td>1</td>
<td>418</td>
<td>18.77</td>
</tr>
<tr>
<td>3</td>
<td>391</td>
<td>25.76</td>
<td>1</td>
<td>.</td>
<td>391</td>
<td>17.56</td>
</tr>
<tr>
<td>Total</td>
<td>1518</td>
<td>100</td>
<td>1100</td>
<td>1127</td>
<td>2227</td>
<td>100</td>
</tr>
</tbody>
</table>

* Percent  
** Number of observations

Setting aside the effect of the reform, one can define prior expectations concerning the changes in the mean number of visits. For the ‘stayer’ population, change should be slightly positive since this sub-population becomes a little older. We expect that recruits will be younger and healthier, whereas the dropouts are likely to be older and less healthy. In both cases, the selection process should tend towards lowering of the population mean. The results are shown in figure A.1 with bootstrapped 95% confidence intervals. Between the two panels, the population experienced a decline from 2.83 to 2.35 visits per person over each three-month observation period. These numbers are similar to the results of Winkelmann (2004), who analysed the change from 1995-1999 for both sexes. For those women who were observed both before and after the reform, the average number of visits rose slightly but did not change significantly. As we are unable to measure changes independently of reform with these data we cannot infer whether this small increase is what one would expect as a result of two years of ageing, or whether it would have been larger but was depressed by the reforms. *Ceteris paribus*, the decomposition suggests that the population decline in visits is a result of statistically-significant selective attrition of individuals who visited physicians at an above-average rate in the pre-reform panel. We cannot expect that a decline of this magnitude will be sustained in the long-run. As expected, the recruits show a lower average number of doctor visits than the total study population. However, the effects of individual change and recruitment have wide confidence intervals and are inconclusive.

We can also use the decomposition technique to examine the impact of selection on covariate statistics. The GSOEP survey used by Winkelmann (2004) included a question on self-reported health that he used as a covariate in his models. Rabe-Hesketh and Skrondal (2008) condensed the 5-point ordinal scale to a binary variable so that the subjective responses ‘very good’, ‘good’ and ‘fair’ were classified as ‘good health’ and coded 0, and ‘poor’ and ‘very poor’ were combined as ‘bad health’ and coded 1. Figure A.2 shows that the proportion experiencing ‘bad health’ fell by about 3% but this was a result of selective attrition of the unhealthy. Winkelmann
Figure A.1.: Decomposition of changes in the mean number of doctor visits from 1996 to 1998 for German women aged 20-60 in full-time employment. The bars represent the population-level change (dark blue), the average individual change (light blue), the change due to dropout (orange) and the change due to new recruits (yellow). The error bars represent 95% confidence intervals.

(2004) found also a declining proportion of people reporting to be in poor health. In our study subjects who were observed twice showed a small, although insignificant, increase in ‘bad health’ as one would expect.

### A.4.2. Example 2: Schizophrenia and Informative Censoring

The data represent 150 subjects taken from a larger, confidential study of drug treatments for schizophrenia and were downloaded from Prof. Peter Diggle’s website (see link in section A.3.1). The outcome variable is a measure of the severity of the symptoms of schizophrenia observed at 0, 1, 2, 4, 6 and 8 weeks, with observed values ranging from 33 to 161. Higher scores relate to more severe symptoms. Subjects were randomised into three treatment regimes of either a placebo or one of two active drugs. For simplicity, we have pooled the data across all three treatments. A larger dataset with the same structure was analysed in Diggle et al. (2007; section 7).

Table A.2 shows the pattern of observations. Only 68 subjects were in observation by week 8 and the attrition was monotone. The dataset is unusual because there is a binary indicator for each individual indicating whether the patient was informatively censored. The decomposition offers an opportunity to assess the sensitivity of this indicator. If the indicator is well defined, individuals are missing completely...
Figure A.2.: Decomposition of changes in the proportion experiencing ‘bad health’ from 1996 to 1998 for German women aged 20-60 in full-time employment. The bars represent the population-level change (dark blue), the average individual change (light blue), the change due to dropout (orange) and the change due to new recruits (yellow). The error bars represent 95% confidence intervals.

Table A.2.: Schizophrenia Symptoms Observation Patterns

<table>
<thead>
<tr>
<th>Follow-up weeks</th>
<th>Cases</th>
<th>Perc.*</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>4</th>
<th>6</th>
<th>8</th>
<th>Obs.**</th>
<th>Perc.*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>68</td>
<td>45.33</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>408</td>
<td>59.56</td>
</tr>
<tr>
<td>2</td>
<td>24</td>
<td>16.00</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>96</td>
<td>14.01</td>
</tr>
<tr>
<td>3</td>
<td>21</td>
<td>14.00</td>
<td>1</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>42</td>
<td>6.13</td>
</tr>
<tr>
<td>4</td>
<td>19</td>
<td>12.67</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>57</td>
<td>8.32</td>
</tr>
<tr>
<td>5</td>
<td>16</td>
<td>10.67</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>80</td>
<td>11.68</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>1.33</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>2</td>
<td>0.29</td>
</tr>
<tr>
<td>Total</td>
<td>150</td>
<td>100.0</td>
<td>150</td>
<td>148</td>
<td>127</td>
<td>108</td>
<td>84</td>
<td>68</td>
<td>685</td>
<td>100.00</td>
</tr>
</tbody>
</table>

* Percent  
** Number of observations

at random and the changes in the mean score contributed by dropouts should be non-significant for the sub-population in which the indicator is zero. The changes should be significant when it is one, but we should note that the indicator is a fixed covariate and therefore may be ‘anticipatory’ until the last observation time. The decomposition is shown in figure A.3. The scores for the total population in the first panel suggest improvement over the whole eight-week period, but selective attrition of high-scoring individuals has a significant effect in masking a deterioration in the later periods for those patients who remain in the study. This dropout effect is not evident in the non-informatively censored group in the centre panel of figure A.3,
Figure A.3.: Decomposition of changes in schizophrenia severity score from one observation time to the next in all data (panel 1), in the non-informatively censored group (panel 2) and in the informatively censored group (panel 3). The label on the x-axis is the follow-up week to which the change was measured. The bars represent the population-level change (dark blue), the average individual change (light blue) and the change due to dropout (orange). The error bars represent 95% confidence intervals.

which suggests that the informative censoring indicator is an effective identification. As expected, the rising scores are particularly evident in those who were regarded as informatively-censored and the dropout effect is more marked (right-hand panel of figure A.3). Diggle et al. (2007) found similar results. They reported an improvement in all treatment groups at the population-level. When a model accounting for the missingness mechanism was applied the individuals getting the placebo did not show an improvement but the two treatment groups did, but to a smaller extend than in the population-level analysis. This suggests that there is an effect of selective dropout, which is strong in the group on placebo and smaller in the treatment groups (Diggle et al. 2007).
Table A.3.: Childhood obesity observation patterns

<table>
<thead>
<tr>
<th>Age groups</th>
<th>Cases</th>
<th>Perc.*</th>
<th>5-6</th>
<th>7-8</th>
<th>9-10</th>
<th>11-12</th>
<th>13-14</th>
<th>15-16</th>
<th>17-18</th>
<th>Obs.**</th>
<th>Perc.*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>460</td>
<td>9.47</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>1380</td>
<td>14.00</td>
</tr>
<tr>
<td>2</td>
<td>448</td>
<td>9.23</td>
<td>.</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>1344</td>
<td>13.64</td>
</tr>
<tr>
<td>3</td>
<td>421</td>
<td>8.67</td>
<td>.</td>
<td>1</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>842</td>
<td>8.54</td>
</tr>
<tr>
<td>4</td>
<td>396</td>
<td>8.15</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>396</td>
<td>4.02</td>
</tr>
<tr>
<td>5</td>
<td>388</td>
<td>7.99</td>
<td>.</td>
<td>.</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>388</td>
<td>3.94</td>
</tr>
<tr>
<td>6</td>
<td>370</td>
<td>7.62</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>1</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>370</td>
<td>3.75</td>
</tr>
<tr>
<td>7</td>
<td>363</td>
<td>7.48</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>.</td>
<td>1089</td>
<td>11.05</td>
</tr>
<tr>
<td>8</td>
<td>287</td>
<td>5.91</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>.</td>
<td>574</td>
<td>5.82</td>
</tr>
<tr>
<td>9</td>
<td>278</td>
<td>5.72</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>834</td>
<td>8.46</td>
</tr>
<tr>
<td>10</td>
<td>250</td>
<td>5.15</td>
<td>.</td>
<td>.</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>500</td>
<td>5.07</td>
</tr>
<tr>
<td>11</td>
<td>238</td>
<td>4.90</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>1</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>476</td>
<td>4.83</td>
</tr>
<tr>
<td>12</td>
<td>221</td>
<td>4.55</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>663</td>
<td>6.73</td>
</tr>
<tr>
<td>13</td>
<td>188</td>
<td>3.87</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>188</td>
<td>1.91</td>
</tr>
<tr>
<td>14</td>
<td>182</td>
<td>3.75</td>
<td>.</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>182</td>
<td>1.85</td>
</tr>
<tr>
<td>15</td>
<td>65</td>
<td>1.34</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>65</td>
<td>0.66</td>
</tr>
<tr>
<td>Other</td>
<td>301</td>
<td>6.20</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>565</td>
<td>5.73</td>
</tr>
<tr>
<td>Total</td>
<td>4856</td>
<td>100.0</td>
<td>353</td>
<td>1378</td>
<td>2245</td>
<td>2268</td>
<td>2033</td>
<td>1215</td>
<td>364</td>
<td>9856</td>
<td>100.00</td>
</tr>
</tbody>
</table>

* Percent
** Number of observations

A.4.3. Example 3: The Muscatine Childhood Obesity Study

The longitudinal data on childhood obesity by age from Woolson and Clarke (1984) has been the subject of a great deal of analysis (e.g. Mahoney et al. 1996; Ekholm and Skinner 1998; Davis et al. 1999, 2001; Birmingham and Fitzmaurice 2002). One of the concerns is that the pattern of substantial non-response is not independent of obesity, the outcome variable. The data come from the Muscatine, Iowa, Coronary Risk Factor Study and children were classified as obese if their weight was more than 110% of the median weight for their sex and height group. Children who feared the embarrassment of a physical examination may not have obtained the required parental consent, or they may have stayed away from school on the assessment day (Ekholm and Skinner 1998). The population is also open to migration which may not be independent of family-level influences on obesity.

The initial panel of 4856 boys and girls in 1977 was classified into five two-year age groups from 5 to 15, and re-examined by cohort in 1979 and 1981. We have reclassified the data by age alone, which leads to repeated measures on incomplete cohorts over ages 5 to 19, but ignores both the age-year interaction and sex differences. Analysing both sexes combined is justified by the analysis of Woolson and Clarke (1984) who did not find a difference between the obesity patterns over age of males and females. Table A.3 shows the 15 most common observational patterns and reveals the complexity of the data.

The decomposition is shown in figure A.4. Population and individual changes in the mean are coordinated, except at ages 11 to 14 when the population shows
an apparently significant decline in obesity. The decomposition suggests that the major part of this decline comes from selection. The dropouts were more likely than average to be obese and those who were not measured at 11 were more likely to be non-obese at 14. Separate analyses by sex suggest that the dropout effect comes mostly from girls whereas the recruitment effect comes from boys. Ekholm and Skinner (1998) and Birmingham and Fitzmaurice (2002) found also that girls who did not respond were more likely to be obese, while males did not show such a pattern. There is an increase in obesity within individuals up to the ages of 9-10, which becomes insignificant for the change to the ages 11-12. From age class 15-16 to age class 17-18 there is a tendency for a decline in the occurrence of obesity, although insignificant due to large confidence intervals. In accordance with this result, Woolson and Clarke (1984) and Birmingham and Fitzmaurice (2002) found a quadratic function in obesity rates with the highest values for the age group 11-15.
A.4.4. Example 4: The Relationship Between Unemployment and Absenteeism in a German Company

Kauermann and Ortlieb (2004) considered the possibility that workers with high absentee rates are more likely to end employment, and that workers who know or expect that their employment will terminate respond with higher absenteeism. Their dataset records the monthly absentee patterns for 150 randomly selected individuals out of a total of 400 employees from one company in Germany between 1988 and 1998. A sample of 100 employees for a longer time period from 1981 to 1998 is available from the website of the Royal Statistical Society (see link in section A.3.1) and used for our analysis. Although the size of the workforce was declining, some recruitment led to non-monotone missingness. We have aggregated the data to a 6 month observational plan. The observational patterns are extremely varied. The most common pattern is employment for the whole observation period, but this group only represents 6% of the cases and the ten most common patterns only cover 37% of the employees. Figure A.5 shows the decomposition of these employment records. The response variable is the proportion of working days that an employee was absent over the preceding 6 month period. The overall absentee rate fluctuated, but it seems that workers who were retained generally increased their absenteeism rates over time. New recruits lowered the rate a little, but it was the attrition of those workers with poor attendance records that offset the upward creep among the retained employees. This result concurs with the findings of Kauermann and Ortlieb (2004).

A.5. Discussion

The decomposition equation that we use in this study is exact, completely general and can be applied to any quantity. We decomposed means, rates and proportions and showed that it is useful in a range of different applications.

Applying the decomposition method to the four examples leads to the following main results:

- The population decline in the doctor visits from before the health care reform in Germany to after is mainly a result of a selective dropout of individuals who went to the physician more often than the average before the introduction of the reform. The decrease in the proportion of individuals experiencing ‘bad health’ is also mainly an effect of unhealthy individuals dropping out.
Figure A.5.: Decomposition of changes in the rate of employee absenteeism in a German company from one 6-month period to the next. The bars represent the population-level change (dark blue), the average individual change (light blue), the change due to dropout (orange) and the change due to new recruits (yellow).

- The scores measuring severity of the symptoms of schizophrenia for the total population suggest improvement over the course of the study, but selective dropout of individuals with severe symptoms masks a tendency for a deterioration of the participants remaining in the study.

- Population and individual changes in the occurrence of obesity are coordinated, except at ages 11 to 14 when the population shows a decline in obesity caused by selective dropout.

- The high absentee rate of employees who subsequently leave the company offsets the trend for decreasing proportion of days absent for the individuals that remain in the company.

If the decomposition shows that there is an effect of selective dropout and therefore data are MNAR, the missingness process cannot be ignored. Complete case analysis, which excludes subjects with incomplete information, weighting of complete cases with the probabilities of missingness, or imputation of missing values are inappro-
appropriate (Schafer and Graham 2002; Fitzmaurice et al. 2004). The most sophisticated analysis for data with MNAR information model the process of the outcome of interest in the study jointly with the missing mechanism. The missing data mechanism is the probability of a response being missing (Fitzmaurice et al. 2004). Possible approaches are selection models, pattern mixture models or correlated random effect models. The selection model jointly models the missingness given the complete data and the distribution of the outcome for the complete data (Little 1995; Schafer and Graham 2002), while the pattern mixture model jointly models the distribution of the outcome given the missingness and the pattern of missingness. This model describes the outcome within groups with different dropout times (Little 1993, 1995; Schafer and Graham 2002). In the correlated random effect models each subject has a random effect that links the model part for the outcome of interest to the model part for the pattern of missingness (Ten Have et al. 2000).

Other methods have been proposed to test for random dropouts (Diggle 1989; Listing and Schlittgen 1998, 2003; Schmitz and Franz 2002). The advantage of the decomposition applied in this study is that it is not only able to provide information on whether the dropout is random but can also test for selective recruitment and can quantify the importance of both processes for the population change.
Bibliography


Bibliography


