

*Sidney De Haan*   
**Research Centre for Arts and Health**

**Singing and Health: A Systematic  
Mapping and Review of Non-Clinical  
Research**

Stephen Clift, Grenville Hancox, Rosalia Staricoff and  
Christine Whitmore



## ***Sidney De Haan Research Centre for Arts and Health***

### Mission:

The Sidney De Haan Research Centre for Arts and Health is committed to researching the contribution of music and other participative arts activities in promoting the well-being and health of individuals and communities.

### Objectives:

- Undertaking scientific research and evaluation on the potential benefits for wellbeing and health of active engagement in music making.
- Documenting and providing the research evidence base for establishing 'Singing on Prescription' for its wellbeing and health benefits
- Working in partnership with health and social care agencies and service users in the South East of England to promote the role of music and arts in healthcare and health promotion
- Contributing to the wider development of the field of Arts and Health research and practice through membership of national and regional networks, publications and educational activities

The Sidney De Haan Research Centre for Arts and Health is part of Canterbury Christ Church University.

<http://www.canterbury.ac.uk/centres/sidney-de-haan-research/>

© Sidney De Haan Research Centre for Arts and Health

Authors: Stephen Clift, Grenville Hancox, Rosalia Staricoff and Christine Whitmore

Published August 2008

ISBN: 978-1-899253-33-3

*Sidney De Haan*   
**Research Centre for Arts and Health**

# **Singing and Health: A Systematic Mapping and Review of Non-Clinical Research**

Stephen Clift, Grenville Hancox, Rosalia Staricoff and  
Christine Whitmore

With Ian Morrison and Matthew Raisbeck

August 2008



<b>CONTENTS</b>	
Tables, Boxes and Figures	5
Appendices	5
Acknowledgements	6
<b>SUMMARY</b>	7
<b>PREFACE</b>	11
<b>1. INTRODUCTION</b>	13
1.1 Aims of the systematic mapping and review	14
<b>2. METHODS</b>	16
2.1 Searching bibliographic databases	16
2.2 Screening of abstracts	16
2.3 Additional search strategies	16
2.4 Quality screening and data extraction	17
2.5 Analysis of sources	17
<b>3. AN OVERVIEW OF THE STUDIES REVIEWED</b>	19
3.1 How the studies were located	19
3.2 Categorisation of studies by design and method	22
3.3 Citation mapping for the studies identified	22
3.4 Designs and the nature of the singing investigated	25
3.5 Participants in the studies under review	31
<b>4. ASSESSMENTS OF RESEARCH ON SINGING AND HEALTH</b>	33
4.1 Qualitative research on the experience and benefits of singing	33
4.2 Surveys of singers using specially constructed questionnaires	44
4.3 Quantitative studies using previously validated psychological measures	56
4.4 Quantitative studies using objective biological measures	66
4.5 Quantitative studies using standardised physical and mental health measures	73
4.6 Quantitative studies using objective physical function measures	93
4.7 Epidemiological studies assessing choral singing as a health determinant	96
<b>5. DISCUSSION</b>	104
5.1 Identifying sources for this review	104
5.2 An initial assessment of the corpus of studies	104
5.3 Benefits associated with singing	105
5.4 Theoretical perspectives on singing, wellbeing and health	112
5.5 Conclusions and recommendations	112
<b>6. REFERENCES</b>	115
<b>APPENDICES</b>	123

<b>Tables</b>	
1. Record of the systematic search process	20
2. Record of the categorisation and pruning process	21
3. Analysis of citation patterns in 35 sources identified for the review	23
4. Design and the character of the singing investigated	26-27
5. Descriptions of singing activity in selected experimental studies	29-30
6. Studies focusing on group singing	31
7. Sex composition of participants in studies of specified groups	32
8. Factor analysis of the music scale items (Hills and Argyle, 1998a)	45
9. Factor analysis of the music scale items (Hills and Argyle, 1998b)	46
10. Responses to 'Singers Emotional Experience Scale' items (Beck et al. 2000)	47
11. Perceived benefits of choral singing – themes identified in the first Clift and Hancox (2001) survey	48
12. The first component emerging from the second Clift and Hancox (2001) survey	50
13. Studies using previously standardised psychological measures	57
14. Results from two studies on singing which measure mood changes	61
15. Studies of the effects of singing using objective biometric measures	67
16. Results from five studies on singing which measure salivary immunoglobulin A	69-70
17. Summary of findings from studies using objective biomarkers other than s-IgA	72
18. Adjusted GHQ-28 and HADS means – intervention and non-intervention conditions (Houston et al. 1998)	74
19. Results from Cohen et al. (2006) on physical health	83-84
20. Results from Cohen et al. (2006) on mental health	85-86
21. Baseline data for selected measures reported in Cohen et al. (2006) and (2007)	89
22. Selected results from Cohen et al. (2007) for physical and mental health	91
23. Median hearing threshold levels in decibels (Steurer et al., 1998)	95
24. Selected data reported by Bygren et al. (1996)	101
25. Choir membership in the national/sex sub-samples in Hyypä and Mäki (2002) (personal communication)	103
<b>Boxes</b>	
1. Three categories of studies identified in the screening of abstracts	17
<b>Figures</b>	
1. Bygren et al. (1996) Estimated survival of people attending cultural events often, occasionally, or rarely	99
<b>Appendices</b>	
1. The systematic search process, February 2007	123
2. Categories for coding of papers identified in searches based on titles and abstracts	133

## ***Acknowledgements***

Our thanks are due to:

Dr. Jeff Brunton, EPPI Centre, Institute of Education, University of London, for assistance in using the EPPI Reviewer software.

Isobel Salisbury, Administrator in the Sidney De Haan Centre, for assisting in the day to day management of the project.

Claire Whiddon, Profile Graphics Ltd, for design work

## **SUMMARY**

Music is one of the defining features of our human nature, and singing is a form of musical participation and expression open to everyone. It is therefore of interest to consider the available evidence on individuals experiences when they sing, and whether singing is beneficial for wellbeing and health.

The aims of this study are: to systematically identify existing published research on singing, wellbeing and health; to map this research in terms of the forms of singing investigated, designs and methods employed and participants involved: to critically appraise this body of research, and where possible synthesise findings to draw general conclusions about the possible benefits of singing for health. The hypothesis underpinning this review is that singing, and particularly group singing, has a positive impact on personal wellbeing and physical health.

### **The search strategy**

Ten bibliographic databases were systematically searched in February 2007 using carefully selected search terms linked to the general themes of 'health', 'music therapy' and 'singing.' Abstracts were screened independently by two researchers and papers categorised by content according to an agreed scheme.

Twenty non-clinical papers of relevance to the review were identified and hard copies obtained. The references cited in these papers were checked for additional sources of interest, and authors were contacted for information on further studies they had published served to identify an additional 14 sources. Professional contacts, physical searching of conference proceedings and general web searches served to add a further 20 sources. In total, therefore, 54 non-clinical reports were identified of potential relevance to the review. The earliest paper identified was published in 1960, with most studies appearing from the late 1990s onwards.

All papers were read independently by two members of the project team using specially devised data extraction and quality screen instruments. Differences of opinion were resolved through discussion. At this stage, a total of 19 studies were excluded for a number of reasons including lack of research data or very limited attention to either singing or health, and duplication of data reported more fully in other sources – leaving 35 papers with a central focus on possible connections between singing, wellbeing and health. The papers identified are highly variable in content, design and quality, but all were included in the review given the relatively recent and under-developed nature of this field of work.

### **Mapping of studies**

Studies were categorised according to their designs and the principal methods employed as follows: qualitative studies (7); surveys employing

specially designed questionnaires (5); quantitative studies using previously validated psychological scales (4); quantitative research using objective biometric measures (8); quantitative research using standardised measures of mental and physical health (3); quantitative research focused on objective performance measures (4), and finally large-scale epidemiological research studies (4). Two qualitative papers relate to the same study but report different data, and three epidemiological papers likewise.

A citation analysis within the corpus of 35 papers revealed a low level of citation. In particular, there was no evidence of increased citation over time. This was taken to reflect the lack of a progressive programme of research in an area of work receiving relatively recent and limited research interest.

The studies varied considerably in the nature of the singing investigated, with some studies concerned with group singing and others with individual singing, some with amateur singers or professional singers or both. The level of detail in the description of singing as an activity also varied considerably from study to study, with some research reports providing surprisingly little information. In terms of samples investigated, studies were again highly variable (e.g. from members of professional opera choruses, large community choirs and a small choir established in a women's prison) and many involved small numbers of participants. In those studies concerned with choir or choral singing a very considerable sex bias was apparent with women on average out-numbering men 3:1.

### **Findings from the review of studies**

Findings from the qualitative and survey studies served to identify recurrent themes regarding the perceived or reported benefits of singing:

- Physical relaxation and release of physical tension
- Emotional release and reduction of feelings of stress
- A sense of happiness, positive mood, joy, elation and feeling high
- A sense of greater personal, emotional and physical wellbeing
- An increased sense of arousal and energy
- Stimulation of cognitive capacities – attention, concentration, memory, learning
- A sense of being absorbed in an activity which draws on multiple capacities of the body and the mind
- A sense of collective bonding through coordinated activity following the same pulse
- The potential for personal contact with others who are like-minded and the development of personal supportive friendships and constructive collaborative relationships
- A sense of contributing to a product which is greater than the sum of its parts
- A sense of personal transcendence beyond mundane and everyday realities, being put in touch with a sense of beauty and something beyond words, which is moving or 'good for the soul'
- An increased sense of self-confidence and self-esteem



- A sense of therapeutic benefit in relation to long-standing psychological and social problems (e.g. depression, a history of abuse, problems with drugs and alcohol, social disadvantage)
- A sense of contributing to the wider community through public performance
- A sense of exercising systems of the body through the physical exertion involved in singing – especially the lungs.
- A sense of disciplining the skeletal-muscular system through the adoption of good posture
- Being engaged in a valued, meaningful, worthwhile activity that gives a sense of purpose and motivation

Two factor analytic studies suggest that a smaller number of dimensions may run through a broad list of benefits. Further research is needed to explore this possibility.

Four studies employed previously devised questionnaires to assess the impact of individual or group singing on 'mood.' Findings are difficult to compare as three different instruments are employed. Surprisingly little clear evidence emerged to support the hypothesis that singing produces a short-term enhancement of positive mood.

Studies employing objective biomarkers of physiological changes occurring when people sing are also highly variable in the character of singing investigated, the characteristics of samples studied and the biomarkers employed. Generally, such studies have chosen measures which are sensitive to arousal and stress. Five studies have used the immune system marker, salivary immunoglobulin A, and four of these report increases in this antibody associated with singing. This is generally interpreted as a positive indicator that singing in some way 'enhances' the immune system, but there is no evidence on how sustained such a change is and whether it actually has any significance for health (e.g. avoidance of upper respiratory infections).

Only two studies have examined the impact of involvement in singing on measures of mental and physical health. In both cases, the focus is on group singing with older and elderly people. Positive findings are reported suggesting that singing can have measurable benefits for mental health and lead to reductions in medication and use of health services. The quality of these studies and the conclusions drawn are open to question however, and further research is needed.

Limited research has been conducted on singing and lung function and appears to show no measurable changes in measures of lung performance even in professional singers (or wind instrument players) when compared with musicians who do use their lungs to make music. One study has investigated the impact of involvement in a professional opera chorus on hearing and presents evidence of significant increase in hearing thresholds across a wide frequency range indicative of hearing damage. This is the only study which suggests that singing might have harmful consequences for wellbeing and health under certain circumstances.

Finally, two large scale epidemiological studies carried out in Sweden and Finland have included 'choir singing' as a possible social determinant of health. Neither study found evidence that participation in singing has an independent influence on health when a range of known determinants of health (e.g. income, education and smoking) were taken into account. Both studies have weaknesses of sampling and measurement, however, and have questionable relevance for assessing the value of singing for health

## **Conclusion**

Given the fact that group singing can be a powerful and moving experience, and given the possibility that singing could contribute to quality of life, wellbeing and even health – it is remarkable that so little research has explored these issues, and disappointing that the research undertaken to date is so limited and so variable in character.

If work on the potential value of group singing for wellbeing and health is to develop, there is a need to create a collaborative and progressive programme of work embracing conceptual clarification / theory building and empirical research. It is hoped that this mapping and reviewing study will help to support the development of such a programme within the community of scholars interested in singing and its potential benefits for health.

## **PREFACE**

### **The Sidney De Haan Research Centre for Arts and Health**

The origins of the Sidney De Haan Research Centre for Arts and Health lie in a small –scale exploratory study carried out by Clift and Hancox (2001) on the perceived benefits of singing held by choral singers. This project was the starting point for a series of cross-Faculty developments in arts and health work within Canterbury Christ Church University. These included: A programme of seminars and workshops entitled Arts, Wellbeing and Health ran for four years (Robinson and Clift, 2002); commissioned research from Arts Council England was undertaken to document community arts and health initiatives across the South East Region (Clift and Vella-Burrows, 2003); research on the value of singing and music for babies before birth and over the first five years of life (Clift, 2002; Clift, Mackenzie and Bushell, 2006), and special issues of the journal *Health Education* (in 2002 & 2005), and the *Journal of the Royal Society for the Promotion of Health* (in 2006) devoted to research on music, arts and health.

Trish Vella-Burrows, a registered nurse and musician, was closely involved in these developments, and helped to bring into focus an interest in the role of music in the care of elderly people and those affected by dementia and Alzheimer’s disease. With her help, and financial support from the Creative Foundation, a research unit for Arts and Health was established in Folkestone in 2003, within a newly established regeneration project based on the creative arts.

Stuart Brown, Chair of Making South East, was a significant supporter at this stage in helping to develop a community singing project for older people, which eventually became the Silver Song Club Project.

With further generous funding from the Roger De Haan Family Trust, the research unit became the Sidney De Haan Research Centre for Arts and Health in 2004. A five-year research strategy was developed and the Centre embarked upon its current programme of research and evaluation focused on the potential health benefits of active participation in music making and singing.

A mission statement was agreed for the Centre, which expressed its commitment to researching ‘the contribution of music and other participative arts activities in promoting the well-being and health of individuals and communities.’

In pursuing this mission, the Centre’s specific objectives are:

- To undertake scientific research and evaluation on the potential benefits for wellbeing and health of active engagement in music making.

- To document and provide the research evidence base for establishing 'Singing on Prescription' for its wellbeing and health benefits.
- To work in partnership with health and social care agencies and service users in the South East to promote the role of music and arts in healthcare and health promotion.
- To contribute to the wider development of the field of Arts and Health research and practice through membership of national and regional networks, publication and educational activities.

The potential value of the arts in healthcare settings and of participation in the arts for individual and community wellbeing has attracted considerable interest in recent years. Throughout the UK, a huge variety of 'arts and health' projects have been established over the last 20 years, and this area has begun to attract serious research attention. Within this broad field of practice and research, the focus of the current work of the Sidney De Haan Research Centre is on the potential value of music-making, and particularly singing, for the wellbeing and health of individuals and communities.

Two substantial strands of research activity on singing and wellbeing began in 2005. Firstly, the Centre had worked in partnership with the charitable company, Sing For Your Life Ltd, to evaluate the benefits for older people of active engagement in community singing groups (the Silver Song Clubs Project) (Bungay, Skingley, Clift and Hancox, 2008). Secondly, it has built on the earlier study by Clift and Hancox (2001), with a cross-national survey of singers in choral societies and choirs in England, Germany and Australia (Clift, Hancox, Morrison, Bärbel, Stewart and Kreutz, 2008).

In planning their initial study, Clift and Hancox (2001) conducted a literature search but found very little previous research on singing and wellbeing. Since then, however, interest in this issue has grown, and at the time of planning the projects described above they were fairly confident, through continual, though non-systematic searching of the academic literature and, more especially through personal contacts with other researchers with an interest in singing and health that they were fully aware of the most important research relevant to their interests. Nevertheless, it was considered essential to go through a systematic process of searching and mapping the existing evidence, and more importantly, a process of critical appraisal, leading to an assessment of current knowledge on the value of singing for wellbeing and health.

The present report is an account of the outcome of this systematic mapping and reviewing process.

# **1. INTRODUCTION**

Before focusing on published scientific research on singing and its effects, it is important to place such research in the context of a wider consideration of the nature and functions of music, the ways in which music is used in everyday life, and the current realities of social participation in music making.

A number of assumptions are taken to underpin our interest in singing and its potential benefits for community and personal wellbeing and health. First and foremost, the human capacity to appreciate and engage with music is a deeply engrained element of our evolutionary heritage (Mithen, 2005), and is underpinned by distinctive features of human neurological structure and function (Levitin, 2006)

Responsiveness to musical sound is apparent in the developing foetus before birth and in pre-mature babies immediately after birth (Cassidy and Standley, 1995; Deliege and Sloboda, 1996). Human language is inherently musical in character, and such musical features as pitch, melody, rhythm, repetition etc. play a central role in the early development of communication between babies and carers prior to the emergence of spoken language (Trevathan, 1999-2000). Singing of lullabies appears to be a culturally universal feature of interaction between mothers and their babies, and serves important functions in regulating the baby's activity and mood, and in establishing the maternal-child bond (Trehub and Trainor, 1998)

All known cultures have musical traditions, and music has served and continues to serve a wide variety of cultural and social functions – some of which may be regarded as socially manipulative (Brown and Volgsten, 2006). A topical example of this is the current use of classical music in underground stations in London in an attempt to reduce antisocial behaviour from young people (Duchen, 2008).

The advent of recorded music, as a relatively recent historical development, has hugely increased the accessibility of an increasingly wide range of diverse styles of music. Particular styles of music can serve as a vehicle for affirming social and personal identities. Music has become a ubiquitous element of visual media – film and television, and is a common feature of public spaces. Such music can powerfully and unconsciously affect cognition, mood and action.

On an individual level, people actively use music as a resource to achieve a variety of ends – particularly in the area of emotion/mood regulation and improving quality of life (Juslin and Sloboda, 2001; Hays and Minichiello, 2005; Coffman, 2002, Juslin and Västfjäll, 2008)). Music has been described as a 'technology of the self' (De Nora, 2000) – a resource that can be used by people for different purposes, including supporting and enhancing a sense of personal wellbeing (Batt-Rawden, DeNora and Ruud 2005; Rohwer and Coffman, 2006). Music does not have effects in and of itself, however, but in accordance with how it is understood and used in social and personal context.

Historically and cross-culturally, music has been regarded as a powerful 'therapeutic' medium, which can help people with significant physical and mental health problems and disability (Gouk, 2000; Horden, 2000). 'Music therapy' continues to be researched throughout the world, and the social and clinical benefits of music have been clearly established (Pavlicevic and Ansdell, 2004; Särkämö, et al., 2008).

Live musical events continue to be in high demand – with contemporary popular music events and festivals attracting huge audiences. Active participation in music making, however, is less widespread in contemporary Western societies, and certainly in the UK, there has been a growing concern that participation in music has been seriously threatened by curriculum changes in schools that have reduced the status of the arts and music. Choral and community singing continues to be the most common form of active musical involvement (Chorus America, 2004), although it tends to be pursued by older sections of the population. A number of studies have explored the value and benefits associated with active participation in music making (Rohwer and Coffman, 2006; Kototsaki and Hallam, 2007).

Active participation in voluntary associations for the purposes of making music has significance in the context of civil society– 'contributing to a stable and inventive society' and maintaining social networks and social capital (Everitt, 1997). Participation in music can take many different forms and occurs to varying degrees. Most people are consumers of music – recorded and live, fewer are actively involved in music making through amateur participation and a small minority pursue committed study of music to become performers (Everitt, 1997).

### **1.1 Aims of the systematic mapping and review**

The broad aims of this project are:

- To employ a systematic search strategy to identify existing published scientific literature on the possible links between participation in singing and wellbeing and health
- To map the characteristics of this literature in terms of the questions it addresses, the conceptual and theoretical frameworks employed, the contexts in which the research was conducted, the individuals and forms of singing investigated, and the methods of data collection and analysis employed
- To systematically and critically appraise the quality of the research undertaken and the evidence presented in order to assess the meaningfulness of conclusions reached
- To attempt to synthesise findings from studies where this is feasible and worthwhile

The hypothesis guiding the review is that:

‘Active engagement in group singing can have a positive effect on individual and community quality of life, wellbeing and health.’

This hypothesis is phrased in a very general way to be maximally inclusive of the studies to be reviewed.

One of the expected outcomes of the review is a greater refinement of an underpinning theoretical framework to guide future research on singing, wellbeing and health.

## **2. METHODS**

### **2.1 Searching bibliographic databases**

The aim of the literature search was to locate published research addressing the possible health benefits associated with active participation in singing. A detailed strategy for the search was devised and carried out by a medical librarian and health information specialist, while on secondment to the Sidney De Haan Research Centre from the East Kent Health Promotion Service (CW). Details of the search are given in Appendix 1. Essentially the search looked for sources related to 'singing', 'health' and 'music therapy' indexed in the following bibliographic databases: Medline, Psychinfo, Cinahl, etc. Searches of these databases were conducted initially in October 2006 and then repeated for a second time in February 2007 to ensure that we had not missed potentially relevant studies published in the course of undertaking the reviewing process.

### **2.2 Screening of abstracts**

Following the searches, abstracts of all identified sources were independently screened by two members of the research team (SC and RS), and coded according to a specially devised category system (See Appendix 2). The results of the two independent screenings were generally very close, and disagreements were readily reconciled by discussion. At this stage, the researchers erred on the side of inclusion rather than exclusion, as in many cases, hard copies of papers had not been obtained.

Papers falling into the first three categories were considered of most relevance to the purposes of the review (See Box 1).

### **2.3 Additional search strategies**

Following the screening process, hard copies of all papers in categories 1-3 were obtained. The references of all papers were then screened to identify any further reports not located by the search of bibliographic databases, which appeared to be relevant to the review. An attempt was also made to contact all principal authors of relevant papers to inform them that the review was underway, and to ask for copies of any additional material they may have published that we had not identified, and information on any current research on singing and health they were involved in or aware of elsewhere.

It was clear at this stage that 'non-clinical' and 'clinical' studies were very different in character. 'Clinical' studies were found concerned with the value of singing in relation to pain relief, aphasia, asthma, Parkinson's disease, dementia, etc. Some of the clinical papers were essentially case studies, focusing on a small number of patients. It was decided, therefore, to focus on 'non-clinical' studies for the purpose of a first mapping and review exercise and consider separately the findings from clinical studies.



### Box 1: Three categories of studies identified in the screening of abstracts

1. Empirical study of any design (quantitative or qualitative), which directly explores the effects of:
  - active participation in singing (solo or group),
  - as a sole activity, or as part of a wider range of activities,
  - on aspects of health, well-being, quality of life or positive biochemical markers,
  - conducted on non-clinical samples.

Highest relevance to the systematic review

2. Empirical evaluation of any design (quantitative or qualitative) which explores the effects of
  - a therapeutic intervention involving active participation in singing (solo or group), as a sole activity,
  - on aspects of a specific condition or more general aspects of health, well-being, quality of life or positive biochemical markers,
  - conducted with clinical samples by a music therapist.

Very relevant to the systematic review

3. Empirical evaluation of any design (quantitative or qualitative) which explores the effects of
  - a therapeutic intervention involving active participation in singing (solo or group), as part of a wider range of activities,
  - on aspects of a specific condition or more general aspects of health, well-being, quality of life or positive biochemical markers,
  - conducted with clinical samples by a music therapist.

Possible relevance to the systematic review

## 2.4 Quality screening and data extraction

Having identified the research papers to be included in the review, details were entered into the EPPI Reviewer software devised by the EPPI Centre, Institute of Education, for managing the process of systematic reviewing. Tools for quality screening and data extraction were specially devised, and applied independently by two members of the research team to each paper (SC, RS, IM and MR). Independent judgements were then compared and any differences were discussed and reconciled.

## 2.5 Analysis of sources

A number of sources were excluded from the review on the grounds that they did not qualify as research papers, had a very limited focus on singing and health or replicated data reported in other sources that were included.

At this point it was clear that the research identified as relevant to our initial guiding hypothesis was highly variable on a large number of parameters (research problem, design, methods, sample, analysis etc.), and also in quality of reporting. In terms of current standards for inclusion of studies in systematic reviews of health research it is likely that a large proportion of the studies identified would not be considered for inclusion on the grounds of methodological or reporting weaknesses.

We decided, however, that this reflected the relatively underdeveloped nature of this field of study, and that there were benefits in mapping the character of published work on singing and health, drawing out a range of issues related to the variations apparent in the nature and quality of the extant literature. This reflective exercise, we believed, would potentially help to identify important issues that need to be addressed in developing an agenda for further research on singing and health in future.

With this in mind, a number of analytical approaches were employed to examine important features of the research identified:

- Firstly, we were interested in how research on connections between 'singing' and aspects of health and wellbeing had developed over time, and whether there was any sense of a progressive direction to the research previously undertaken. This was done by ordering the research papers chronologically and undertaking a citation mapping to assess the degree to which researchers were aware of previously conducted studies.
- Secondly, we were interested in the research problems identified by researchers, the designs employed, the character of the singing investigated, and the kinds of data gathered. Sources were categorised in these terms. This provided a useful structure for critically reviewing the studies.
- Thirdly, we were interested in critically assessing the findings reported by studies, identifying common themes regarding possible health benefits, considering the strength of the influence of singing on health and wellbeing established by the research and determining the extent to which findings had been independently replicated.
- And finally, we were interested in drawing together the theoretical frameworks and models employed or devised by researchers to guide their studies and make sense of their findings.

### **3. AN OVERVIEW OF THE STUDIES REVIEWED**

#### **3.1 How the studies were located**

In total, 54 sources were located following the search strategies outlined above. These are given in first part of the References.

Table 1 reports details on how these sources were located. Column one reports the bibliographic source employed and column two the number of sources identified. Databases were searched in the order indicated in the table, and duplicate sources were ignored. Column three gives the papers identified on the basis of screening titles and abstracts as potentially relevant to a review of non-clinical research. Column four gives additional sources identified from the references included in these papers, and column five indicates whether any additional sources were identified through personal communication with the principal authors of papers identified ('No follow up indicates that communication via email or letter was not successful).

The final column gives the total number of sources identified arising out of the initial database search. As indicated, 20 out of the 53 sources identified were located via a number of different routes, and sometimes in an entirely serendipitous way.

#### **3.2 Categorisation of studies by design and method**

All of the these papers were read independently by two members of the research team following the procedures outlined above and at this point decisions were made on 1) whether a paper should be included in the mapping and review process and 2) the basic character of the research undertaken. The results of this process are given in Table 2. The organisation of research studies from category 6 to category 12 could be read a hierarchy of evidence, with the data provided by studies becoming progressively more quantitative and controlled. However, this reading should be resisted, as the more controlled and quantitative studies do not necessarily provide 'harder' or more illuminating evidence on the possible links between participation in singing and health. It is more realistic and fruitful to think of a typology of research rather than a hierarchy (Peticrew and Roberts, 2005) with different kinds of studies addressing different questions and providing appropriately different kinds of evidence. Ideally, one would hope to find evidence of a consistent and complementary kind across of range of differently designed studies, all of which have a related focus on singing and its effects.

At this point 19 papers were removed from further consideration on the grounds that they did not report empirical research findings, duplicated data reported elsewhere, did not have a clear focus on health, were primarily concerned with singing and made very minor references to health, or were concerned in a general way with the arts in relation to health and made very minor references to singing. In total, therefore, 34 studies were considered of interest for more detailed consideration in this review.

Table 1: Record of the systematic search process					
Database	Feb 2007 No. sources identified	Papers judged to be relevant to a review of non-clinical research	Papers cited	Follow-up with author	Total no.
Medline	82	Kreutz (2004) Grape (2003) Clift (2001)  Valentine (2001) Brygen (1996) Wise (1992)	No new papers No new papers Schorr-Lesnick (1985) Rider (1991) Hills (1998a,b) Anshel (1988) No new papers No new papers	Follow up - no new papers Follow up, work in progress N/A  No follow up Follow up – Konlaan (2000), Johansson (2001) Follow up - no new papers	13
Embase	74	No new papers			
Cinahl	39	Sandgren (2002) Cusicanqui (2004)	No new papers No new papers	Follow up, work in progress Follow up - no new papers	2
Psycinfo	38	Faulkner (2006) Bailey (2002b)  Bailey (2003b) Bailey (2005)  Unwin (2002) Beck (2000) Houston (1998) <sup>2</sup> Montemurro (1996)	Faulkner (2004) Adams (2000) Patteson (2000) No new papers Pitts (2004)  No new papers No new papers No new papers No new papers	Follow up - no new papers <u>Follow up with Jane Davidson – Bailey (2002a, 2003a) and work on-going - publication to appear in Edwards (ed) (2008)</u> <sup>1</sup> No follow up Follow up - Beck (2006) No follow up No follow up	15
ERIC	38	Michaelos (2005) Silber (2005) Joyce (1996)	No new papers No new papers No new papers	Follow up - no new papers Follow up - no new papers Follow up - no new papers	3
ASSIA	69	No new papers			0
RILM	196	No new papers			0
Cochrane Library	85	No non-clinical research papers			0
Music Therapy World	C 400	Steurer (1998)			1
Google Scholar	First 100 of C 40,000 records examined	No new papers			0
Additional sources <sup>3</sup>	Cohen (2006, 2007), Windsor (2005), Chorus (2003), Kreutz (2003), Silvey (2002), Hillman (2002), Kuhn (2002) Ashley (2002), Smithrim (1998), Gould (1973), Heller (1960), Durrant (1998), Richards (2003), Watanabe (2005), Kennedy (2002), Galati (2006) Louhivuori (2005, 2006), Hyyppa (2001)				20
					<b>54</b>

1. Edwards, J. (ed.)(2008) Music and Health, Cambridge: Cambridge Scholars.

2. Houston et al. (1998) study conducted in a care setting – but not a music therapy intervention for a specified health condition

3. Additional sources – already known from personal contacts, citations in cited papers, identified from general web based searches for music and health, or from following up leads from papers identified through the search process (e.g. papers published in 'The Phenomenon of Singing' Volumes 1, 2 and 3), which had some relevance but not referenced in any sources.

Table 2: Record of the categorisation and pruning process		
1. Removal of all papers that are <u>not</u> research reports (they may well make reference to the value of singing, but claims are not backed up by formal research or the presentation of data)	Cusicanqui (2004), Montemurro (1996)	2
2. Removal of research papers which simply duplicate data (e.g. conference proceedings reports of studies that are published elsewhere)	Bailey (2002a), Bailey (2003b) Kreutz (2003). Lourivouri (2006)	4
3. Removal of research papers which are not concerned with the effects of singing on wellbeing/health or related concepts	Anshel (1988)	1
4. Research papers/reports on wellbeing and health in relation to the arts in general - not solely concerned with singing, but which make minor reference to it	Michalos (2005), Windsor (2005)	2
5. Research papers on singing which make <u>limited</u> references to wellbeing and health (they may report some data or simply make general comments based on the study undertaken)	Adams (2000), Chorus (2003), Durrant (1998), Faulkner (2004, 2006), Kennedy (2002), Pitts (2004), Richards (2003), Silvey (2002), Watanabe (2005)	10
6. Qualitative research papers that report on singers' experience of singing and its effects on wellbeing and health	Ashley (2002), Bailey (2002b, 2005), Joyce (1996), Silber (2005), Patteson (2000), Smithrim (1998)	7
7. Research studies which report numerical data using specially devised measures of singers' experiences of singing, and/or its effects on wellbeing and health (and may gather qualitative data – category 6)	Bailey (2003a), Clift (2001), Hillman (2002), Lourivouri (2005), Sandgren (2002)	5
8. Quantitative research studies that use previously validated psychological measures e.g. of emotion or mood (and may use specially devised measures – category 7)	Unwin (2002), Wise (1992), Hills (1998a,1998b)	4
9. Quantitative studies that use objective measures of biometric variables (and may use specially devised or previously validated measures - categories 7 and 8)	Beck (2000, 2006), Galati (2006), Grape (2003), Kreutz (2004), Kuhn (2002), Rider (1991), Valentine (2001)	8
10. Quantitative studies that use standardised physical and mental health measures	Cohen (2006, 2007), Houston (1998)	3
11. Quantitative studies that use objective physical performance measures (i.e. lung function, hearing)	Heller (1960), Gould (1973), Schorr-Lesnack (1985), Steurer (1998)	4
12. Epidemiological research on determinants of health which include reference to cultural or leisure factors, including singing in choirs	Hyypa (2001), Bygren (1996), Konlaan (2000), Johansson (2001)	4
		54

### 3.3 Citation mapping for the studies identified

As indicated above, it was considered of interest to examine how research on 'singing' and aspects of health and wellbeing had developed over time, and whether there was any sense of a progressive direction to the research previously undertaken. This was achieved in Table 3. by ordering sources by date (and within date alphabetically) (columns one and two), and looking at the extent to which previously published papers were cited. Reading across the rows from each source gives the number of previously published papers cited in each source, and reading down the columns from each source gives the number of times that source is cited.

If a progressive process of taking account of previous research and building upon it, has taken place, one would expect to see in the final column, a general increase in the number of citations of previously published papers over time. Similarly, one would expect to see the last row, a fall in number of times a paper is cited, as the scope for citation clearly diminishes.

In fact, the results of this analysis show a very low degree of citation with the set of sources identified.

The most cited and citing sources are indicated with blue highlighting in this table:

Most cited: Bygren 1996, Beck 2000, Clift 2001 and Kreutz 2004

Most citing: Clift 2000, Bailey 2002b, Bailey 2003a, Kreutz 2004, Bailey 2005 and Beck 2006

This listing identifies a core of researchers who are interested in the potential value of singing and health, and who are mutually aware of one another's research activity.

Table 3: Analysis of citation patterns in 35 sources identified for the review

Date of publication	Column: First author of papers identified	Rows: Citation of papers within the set identified	Heller 1960	Gould 1973	Schorr-Lesnick	Rider 1991	Wise 1992	Bygren 1996	Joyce 1996	Hills 1998a	Hills 1998b	Houston 1998	Smithrim 1998	Steurer 1998	Beck 2000	Konlaan, 2000	Patteson 2000	Cliff 2001	Hyyppa 2001	Johansson 2001	Valentine 2001	Ashley 2002	Kuhn 2002	Bailey 2002b	Hillman 2002	Sangren 2002	Unwin 2002	Bailey 2003a	Grape 2003	Kreutz 2004	Bailey 2005	Louhivuori 2005	Silber 2005	Beck 2006	Cohen 2006	Galati 2006	Cohen 2007	No. of papers cited				
1960	Heller		X																																						0	
1973	Gould			X																																						1
1985	Schorr-Lesnick		X		X																																					2
1991	Rider					X																																			0	
1992	Wise						X																																		0	
1996	Bygren							X																																	0	
1996	Joyce								X																																0	
1998a	Hills									X																															1	
1998b	Hills										X																														0	
1998	Houston											X																													0	
1998	Smithrim												X																												1	
1998	Steurer													X																											0	
2000	Beck														X																										1	
2000	Konlaan															X																									1	
2000	Patteson																X																								2	
2001	Cliff																	X																							2	
2001	Cliff																		X																						5	
2001	Hyyppa																			X																					1	
2001	Johansson																				X																				0	
2001	Valentine																					X																			2	
2001	Valentine																						X																		1	
2002	Ashley																							X																	0	
2002	Kuhn																								X																0	
2002b	Bailey																									X															3	
2002	Hillman																										X														1	
2002	Sandgren																											X													0	
2002	Unwin																												X												0	
2003a	Bailey																													X											3	
2003	Grape																														X										1	
2004	Kreutz																															X									5	
2005	Bailey																															X									3	
2005	Louhivuori																																X								1	
2005	Silber																																	X							1	
2006	Beck																																		X						6	
2006	Cohen																																			X				0		
2006	Galati																																				X				0	
2007	Cohen																																				X				1	
	Times cited																																									0

+ Reference is to an MSc dissertation by Joyce (1993) on which this paper is based.

This analysis also reveals other findings of interest:

- The three earliest papers focus on the question of whether singing is associated with measurable improvements in lung function. This is a research question of considerable interest, but clearly and surprisingly, no further research appears to have been conducted on this topic since the report by Schorr-Lesnack et al. in 1985.
- Nine studies were published during the 1990, and citations within this set are virtually non-existent, and similarly few are cited subsequently, with the obvious exception of Bygren et al. (1996). This paper reports a significant and large-scale longitudinal survey of cultural participation in relation to survival among the Swedish population. Singing in choirs and making music was one aspect of cultural participation considered in this study. A further follow up phase of the Bygren et al. study is report by Konlann et al. (2000), but interesting the dimension of 'singing in choirs' is not specifically highlighted, being subsumed under the more general label of 'making music'. The later report is missed by Clift and Hancox (2001), Hyyppa and Maki (2001) and Valentine and Evans (2001), most likely because of the close dates of publication, and like Bygren et al. is not cited by any further report after 2001.
- Something of a turning point occurs in 2000 with the publication of a study by Beck et al. (2000) focused on psycho-physiological dimensions of choral rehearsal and performance (immunoglobulin A and cortisol secretion). This paper interestingly cites none of the earlier studies identified, nor is it cited by studies published shortly after by Clift and Hancox (2001), Valentine and Evans (2001), as all three studies were underway and in the process of publication over the same period of time. Surprisingly, Beck et al. (2000) is not cited by Kuhn (2002) despite the clear overlap in focus on immunoglobulin secretion in saliva in response to singing, but again, there may be an issue of timing of the research and length of time it takes to get results into print. The paper by Beck et al. (2000) is subsequently cited by Bailey and Davidson (2002b, 2003a), Kreutz et al. (2004) and Beck et al. (2006). Both Kreutz et al. (2004) and Beck (2006) fail to cite Rider et al. (1991) and Kuhn (2002) even both earlier studies examine immunoglobulin A levels in singers.
- The study by Clift and Hancox (2001) emerges as the most cited study in the set of sources identified, being cited by same researchers who cite Beck et al. (2000), but also by Hillman (2002) and Ashley (2002) (studies cited by no other source subsequently).
- The work of Bailey and Davidson (2002b, 2003a, 2005) deserves special mention as they contribute three important papers to the set of sources under consideration. They cite Clift and Hancox (2001) Beck et al. (2000) and Kreutz (2004) and also Patteson (2000) (a study cited by no other source subsequently). Despite the significance of Bailey and Davidson's 2002b and 2005 papers, which report comprehensively on



Bailey's PhD research with choirs established for homeless and disadvantaged groups, and present a very significant theoretical model of the benefits associated with group singing, neither of them is subsequently cited by other authors.

- The remaining papers published in 2005 and beyond are an interesting group. Silber (2005), which reports an ethnographic study of the value of a singing group established in a women's prison in Israel. None of the previously published papers in the set are referenced, and it is especially surprising to see that no mention is made of the work of Bailey and Davidson and earlier studies focused on the value of singing for women (e.g. Joyce, 1996; Smithrim, 1998). Louhivuori et al. (2006) report a significant study of choral singing and social capital with data from choirs in Europe and Africa, but again make no reference to previous research identified in this search. The same is true of Cohen et al. (2006, 2007) who report what is probably the most significant study on singing and its health benefits published to date. Galati et al. (2006) report a rather specialised study and somewhat artificial study exploring the emotional and physiological impact of singing on individual singers, and it is perhaps less surprising to see no references made to previous literature. Finally, a second study by Beck and colleagues (2007) refers to no less than six previous studies identified by the search, which undoubtedly reflects citations to his earlier study (Beck et al. 2000) by a core of scholars interested in choral singing and health.

### **3.4 Designs and the nature of the singing investigated**

Table 4 elaborates on the framework presented in Table 2 and presents information on the designs of studies identified and the character of the singing investigated.

Clearly, as one moves from qualitative studies listed initially, through structured surveys of choirs, quasi-experimental and experimental studies, focused physical performance studies and finally large scale epidemiological surveys, so the nature of the designs, the samples, the methods, the data gathered and the analytical techniques, all change. It is striking also that is very little evidence of any replication and planned extension of previous studies, with the exception is the research by Kreutz et al. (2004), which very specifically takes the Beck et al. (2000) study as its starting point. As noted earlier, we do not regard this table as presenting a hierarchy of evidence, as studies adopting different designs and approaches to data gathering are concerned with different kind so problems and provide different kinds of evidence and insight.

Table 4: Design and the character of the singing investigated

Qualitative studies of singers' experiences			
Date	First author	Design/Method	Nature of singing group / activity investigated
1996	Joyce	Qualitative study, interviews	Women speaking about the process and effects of singing with other women
1998	Smithrim	Qualitative study, interviews	Reflections by women on the place of singing in their lives – both individually and in groups
2000	Patteson	Qualitative study, interviews	Reflections by women on the value of individual singing lessons
2002	Ashley	Ethnographic study, observation, diary keeping, individual and group interviews	Boy choristers singing in an established church choir
2002b/2005	Bailey	Qualitative study, interviews, focus group discussion	Disadvantaged and advantaged groups singing in established amateur choirs
2005	Silber	Ethnographic study, observation, interviews	Small choir set up in a women's prison, meeting once a week for a period of eight months.
Studies of singers' experiences with specially designed questionnaires			
Date	First author	Design/Method	Nature of singing group / activity investigated
2001	Clift	Questionnaire survey of experiences associated with choral singing	Membership of an established choral society
2002	Hillman	Questionnaire survey of experiences associated with singing	Membership of an established community choir
2002	Sandgren	Qualitative and quantitative study, interviews and structured assessments of health problems	Professional opera singers
2003a	Bailey	Questionnaire survey of the same participants experiences associated with three levels of musical participation	Membership of established amateur choirs
2005	Louhivuori	Cross-national, comparative questionnaire survey of experiences associated with choral singing	Membership of established amateur choirs
Quantitative research studies that use previously validated psychological measures			
Date	First author	Design/Method	Nature of singing group / activity investigated
1992	Wise	Comparative survey of choral singers and non-singers in a retirement community, with standardised measures	Membership of a community choral singing group
1998a	Hills	Questionnaire survey of experiences associated with leisure activities, including music	Membership of established music groups / choirs
1998b	Hills	Questionnaire survey of experiences associated with leisure activities , including music	Membership of established music groups / choirs
2002	Unwin	Experiment, separate intervention (singing) and control (listening) groups, pre/post assessment of mood	Singing group set up for the study, group singing for half an hour
Quantitative studies that use objective measures of biometric variables			
Date	First author	Design/Method	Nature of singing group / activity investigated
1991	Rider	Quasi-experiment, three conditions plus control, same participants in all conditions, pre/post assessments of mood and physiological variables	Participants with musical training singing familiar songs for ten minutes
2000	Beck	Quasi-experiment, two conditions, same participants in both, pre/post assessment of mood and physiological variables	Established professional choir rehearsing and performing Beethoven's <i>Missa Solemnis</i>
2001	Valentine	Quasi-experiment, three separate groups compared, pre/post assessment of mood and physiological variables	Individual singers involved in a single lesson, choral singers involved in a rehearsal and swimmers at a local leisure centre

Table 4 continued: Design and the character of the singing investigated			
2002	Kuhn	Experiment, two separate intervention groups and one control, pre/post assessment of physiological variables	Small singing group set up for study, single session of singing for half an hour
2003	Grape	Quasi-experiment, two separate groups compared, pre-post assessment of mood and physiological variables	Individual amateur and professional singers assessed before and after a single singing lesson
2004	Kreutz	Quasi-experiment, two conditions (singing and listening), same participants in both, pre/post assessment of mood and physiological variables	Established amateur choral society rehearsing Mozart's <i>Requiem</i> for one hour
2006	Beck	Quasi-experiment, repeated pre/post assessment of mood and physiological variables	Individual student singers involved in lessons and performances over a period of 10 weeks
2006	Galati	Quasi-experiment, three conditions, singers acting as own controls, pre/post assessment of mood and physiological variables	Professional and semi-professional individual singers, singing six 40 second melodies using nonsense syllables
Quantitative studies that use standardised physical and mental health measures			
Date	First author	Design/Method	Nature of singing group / activity investigated
1998	Houston	Experiment, singing intervention and control groups, pre/post assessment of health variables	Singing group set up for study, one hour weekly for four weeks
2006/ 2007	Cohen	Quasi-experiment, singing intervention and control, pre/post assessment of health with standardised and objective measures over one (2006) and two(2007) years	Choir set up for the study, rehearsing and performing over 30 weeks in year one and year two
Quantitative studies that use objective physical performance measures			
Date	First author	Design/Method	Nature of singing group / activity investigated
1960	Heller	Comparative study of professional singers and a group with no professional vocal training	Professional singers
1973	Gould	Comparative study of professional singers, students singers and non-musicians	Professional and student singers
1985	Schorr-L	Comparative study of singers, wind and string/percussion instrumentalists	Singers with the New York City Opera and Choristers' Union
1998	Steurer	Assessment of hearing threshold levels	Singers with the choir of the Vienna State Opera
Epidemiological research on determinants of health			
Date	First author	Design/Method	Nature of singing group / activity investigated
1996/ 2000/ 2001	Bygren / Konlaan/ Johansson	Large scale population survey with longitudinal follow up assessing participation in leisure activities and links with survival	Membership of established music groups / choirs
2001	Hyypa	Large scale cross-sectional survey of two language populations, assessment of social and health variables	Membership of established choirs

In addition to all the variations in design and methods, studies also vary very markedly in the character of the singing investigated. This issue is actually of much greater significance in assessing the corpus of studies identified here than are matters of design and data collection, as it concerns the core issue of what the research studies are actually concerned with, when they investigate singing and its possible effects and benefits.

Once attention is focused specifically on the character of singing investigated and the extent to which researchers describe or more importantly, neglect to describe the activity they are researching, three issues come into focus: the diversity of the singing activity investigated, the paucity of detail provided in most studies, and the lack of any agreed theoretical or even typological framework within which to locate the activity focused on in any particular study.

It is not intended to undertake a detailed analysis of the information on the character of the singing provided in all of the papers identified (primarily because often little description is provided), but it is of interest to consider the way in which quasi-experimental / experimental studies, describe the singing activity investigated. In these studies, singing is regarded as an 'intervention', or 'treatment' or 'experimental condition' or 'independent variable' and conceived as such, it is clearly important to consider carefully the descriptions provided.

Table 5 below offers selected examples from papers which provide relatively detailed accounts of the singing studied and it is immediately apparent that the nature of singing activity involved varies enormously. Studies range from assessing the impact of involvement in a community choir over a period of a year, to assessing the impact of singing newly composed melodies with nonsense syllables over a period of two minutes! In addition, the descriptions vary in the kind of information given on such issues as the nature of the repertoire, the context in which the singing took place, whether there was accompaniment, the role of a conductor or leader, whether the singers were standing or sitting, and so on.

This exercise points strongly to the need for a formal analysis of the nature of singing and how it can vary – both with respect to the parameters of physical / physiological mechanisms operating in the singer's body, and in relation to the social arrangements under which an individual sings or comes together with others in a group setting to sing.

Table 5: Descriptions of singing activity in selected experimental studies

	Experimental studies with existing choirs
Beck 2000	'The singers of the Pacific Chorale, a professional performing group (N=180), constituted the initial pool from which a sample of singers was selected. (p.93) 'The principal independent variable was choral singing, and the dependent variables were measures of S-IgA and cortisol concentrations in saliva. A naturalistic study was conducted by using pre-singing and post-singing measures of S-IgA and cortisol levels during two rehearsals and one performance that had been previously scheduled by the professional chorale. In effect the rehearsals and performance served as natural treatment conditions. (p. 92) Saliva samples were collected immediately before (7pm) and after (9.30pm) an early rehearsal and a late rehearsal and before (8pm) and after (9.30pm) a stage performance of Beethoven's choral masterwork, the <i>Missa Solemnis</i> . (p.93)
Kreutz 2004	Members of an amateur choir. '... the experimental conditions for this study were realized in two sessions at the same location in the rehearsal room of a church at the regular time of that choir's rehearsal between 6 and 7 pm.'(p.626) 'The singing condition was initiated by a 10-min warm-up phase, in which various breathing, stretching and vocalising exercises were performed. For the rest of the session, sections and pieces from Mozart's <i>Requiem</i> were rehearsed, and instructions by the conductor were given to the choir. Participants stood during the warm-up, whereas they remained seated for the rest of the time. Times of interruption by the conductor were measured and approximated 10 mins of rehearsal time.' (p.627)
	Experimental studies with specially established singing groups
Houston 1998	'The session consisted of the researchers singing and dancing in a comical fashion to the backing of a pre-recorded cassette played on a portable cassette player (...). The participants were encouraged to join in with the singing and provided with coster hats and jackets if they so required. The session took place in the communal day areas of each care setting, and lasted for one hour.' (pp.329-30) The intervention ran over four weeks.
Kuhn 2002	'Thirty-three undergraduate students at Willamette University ... volunteered to participate in the study. (...) Participants ... were randomly assigned to one of three groups: active, passive, or control.' (p.33) 'Each group consisted of 5-6 participants at a time and ran for 30 minutes. The active groups (n=11) consisted of a music making session facilitated by a music therapy student. Subjects participated in call and response drumming and singing activities as well as short improvisations on a pentatonic scale using tone bars and drums.' (p.34)
Unwin et al. 2002	'The singing programme consisted of two vocal exercises and five songs. Exercise One was an arpeggio [musical details given]. Exercise Two required participants to sing the numbers 1 to 8, staccato (detached) and allegro (at a lively tempo) [musical details given]. The five songs, from Paraguay, Argentina, Africa and the Caribbean, were simple in structure, melody, tonality and rhythm [musical details and justification of choice given]. The songs (...) provided a variety of melodic and rhythmic features, had non-English lyrics and were unfamiliar. The choice of non-English lyrics was, in part, influenced by research indicating that the influence of lyric content may override musical accompaniment in altering the mood of

	<p>listeners. (...) The level of complexity was also minimised by choosing songs with repetitive lyrics, so as to make it easier for the volunteer singers to rapidly become familiar with the task. The first author, an experienced singing teacher, spoke the words clearly and energetically and articulated the words in rhythm, encouraging the singing group to repeat in like manner using an overhead of the music and words as an aid. She then sang the song and invited the singing group to follow, continuing in this manner until the group was familiar with the song (usually two or three times). A similar procedure was undertaken with the remaining four songs. When the teaching of the five songs was completed, the overheads were again displayed one at a time, and the group sang the five songs without further instruction. All songs were sung more than once without pausing between each repetition, some as many as six times, with dynamic and tempo variations directed by the instructor. All songs were accompanied on the piano by the second author.' (pp.78-79)</p>
Cohen 2006	<p>'The notice for the intervention group differed only in that it sought singers for a chorale; no singing experience was required, and the study's purpose was to explore the impact of this activity on general health and mental health as well as involvement in overall individual and community activities of older adults living in the community.' (p.728) 'The intervention consisted of participating in a professionally conducted chorale in which there were weekly rehearsals for 30 weeks as well as public performances several times during the intervention period.'(p.729) 'The chorale was 'facilitated through the Levine School of Music.' (p.728)</p>
	<p>Experimental studies with individual singers</p>
Rider 1991	<p>'Singing (S). Subjects in this condition sang one or two verses of ten familiar songs while they listened to piano accompaniments. (These songs were: "You Are My Sunshine," "He's Got the Whole World in His Hands", "My Country Tis of Thee," "Amazing Grace", "Kum Ba Yah," "When the Saints Go Marching In", "Jingle Bells," "Michael Row the Boat Ashore," "Oh, Susanna," and "Silent Night." Songs were played in comfortable keys for singing; if subjects could not remember verses, they were told to repeat familiar verses. Familiar songs were used to help minimise possible left-hemisphere involvement.' (p. 77) The singing condition lasted for ten minutes.</p>
Dario 2006	<p>'Ten 40-second long melodies were written by a group of five composers according to the rules of western music. Five melodies were composed to have a happiness value, and five to have a sadness value. Happy melodies were written with a fast tempo and major mode, sad melodies were written with a slow tempo and minor mode. Then three musicians chose three happy melodies and three sad ones in order to form the set of experimental stimuli with happy and sad conditions highly differentiated.' (p.1886) 'Two weeks before the experiment the singers received the six melodies and they were asked to learn the melodies using non-sense syllables.' (p.1887) In the experiment, singers sang two blocks of songs – happy and sad – for a total of two minutes for each block.</p>

### 3.5 Participants in the studies under review

The studies in the corpus identified vary considerably in the characteristics of the participants investigated. Some studies focus on amateur singers and some on professional; some are concerned with choral singing and others with amateur singing. In some cases, members of existing groups are investigated and in others, groups are established specifically for research purposes. As the main interest in this review is to explore the idea that amateur group singing may have benefits for wellbeing and health, Table 6 lists those studies which are specifically concerned with members of existing amateur singing groups, professional singing groups and singing groups established for the purpose of research. These studies will be given greater attention in the analysis and discussion which follows.

Table 6: Studies focusing on group singing		
Amateur group singers	Professional group singers	Specially established groups
Wise et al. (1992) Hills and Argyle (1998a,b) Clift and Hancox (2001) Valentine (2001) Ashley (2002) Hilman (2002) Bailey and Davidson (2002b, 2003a, 2005) Kreutz et al. (2004) Silber (2005) Lourivouri et al.(2005)	Beck et al. (2000) Steurer et al. (1998)	Kuhn (2002) Houston (1998) Unwin (2002) Cohen (2006, 2007)

A further issue that should be highlighted is the fact that in many of the studies considered in this review, women out-number men quite substantially. As a principal concern of the review is to explore the benefit of group or choral singing for wellbeing and health, it is of particular interest to consider the proportions of men and women involved in studies of existing choral societies or choirs, and in singing groups specifically set up for research.

Table 7 reports the number of males and females participating in such studies. Wise et al. (1992), Valentine and Evans (2001) and Louhivouri et al. (2005) are omitted from Table 7, as the sex compositions of their samples are not specified. Hills and Argyle (1998a) simply state that the panel sample involved in their studies consisted of 37% men.

Table 7: Sex composition of participants in studies of specified groups					
Study	Total	Women		Men	
	N	N	%	N	%
Steurer, 1998	62	30	48.4	32	51.6
Beck 2000	41	23	56.1	18	43.9
Clift, 2001	91	74	81.3	17	18.7
Hillman, 2002	75	60	80.0	15	20.0
Bailey 2003	124	97	78.2	27	21.8
Kreutz 2004a	31	23	74.2	8	25.8
Bailey 2005 <sup>1</sup>	23	10	43.5	13	56/5
Houston 1998 <sup>2</sup>	61	44	72.1	17	27.9
Ashley, 2002	18	0	0.0	18	100.0
Kuhn, 2002 <sup>2</sup>	33	28	84.8	5	15.2
Unwin, 2002 <sup>2</sup>	107	84	78.5	23	21.5
Silber, 2005	7	7	100.0	0	0.0
Cohen 2006 - Chorale <sup>3</sup>	90	70	78.0	20	22.0
	763	550	72.1	213	27.9

1. Number of men and women in choir 2 not explicitly stated – numbers judged from interview quotations reported
2. Breakdown by sex given for the total sample and not separately for intervention and comparison conditions
3. Number and sex breakdown at pre-test for the chorale intervention group



## **4. assessments of research on singing and health**

### **4.1 Qualitative research on the experience and benefits of singing**

A number of qualitative studies have been undertaken which explore personal experiences and perceived benefit of singing. These studies vary in their methods and participants, with most relying on interviews and discussions with small number of individuals. An interesting feature of this set of studies is the strong focus on a single sex. Ashley (2002) and Bailey and Davidson (2002b) report studies of choir boys and homeless men. Joyce (1996), Patteson (2000) and Smithrim (1998) focus on the experiences of women, and Silber (2005) describes the experience of setting up a choir in a women's prison. Bailey and Davidson (2005) report two further studies involving both men and women. All of these studies are small in scale and employ interviews or discussion groups as the primary means of data collection. Studies by Joyce, Patteson, Smithrim and Bailey and Davidson are retrospective in character, whereas Ashley and Silber report findings from longer term ethnographic studies. A further thread which runs through a number of these studies is a focus on marginalised and disadvantaged individuals (Bailey and Davidson, 2002b, 2005) and the value of singing in promoting personal growth and healing (Joyce, 1996).

In reviewing this material, the three earliest studies by can usefully be considered together, followed by individual attention to the work of Ashley, Bailey and Davidson and Silber.

Two papers report a combination of qualitative and quantitative questionnaire data on experiences of singing. Clift and Hancox (2001) report two surveys of choral singers, one qualitatively focused and one more quantitative. Sandgren (2002) reports a qualitative and quantitative study of professional opera singers. These studies will be considered in the following section of this review.

#### **4.1.1 Early studies of women and singing**

The titles of these early studies immediately give some sense of their character:

*Turn off the radio and sing for your lives! Women, singing and experiential education* (Joyce)

*Still singing for our lives: singing in the everyday lives of women through this century* (Smithrim)

*Singing a woman's life: how singing lessons transformed the lives of nine women* (Patteson)

The emphasis is clearly on highlighting the power of singing in the lives of women. The studies are essentially exercises in personal testimony and

advocacy for the power of singing in terms of personal sustenance and transformation. To get a sense of this, it is useful to present the authors' descriptions of their purposes and the participants involved in their studies.

Joyce (1996) explains the nature of her work as follows:

'This chapter is a distillation of a research project (...) in which I asked seven women to speak about the process of singing and its effects on learning, healing and transformation. All of the women who participated in the study sing with other women in a variety of contexts, with the common denominator being that they are engaged in some process of change or growth in their lives. As a singer and facilitator of women singing in the context of a wilderness program for survivors of chronic abuse, I included myself in this participatory project.' (1996: 253-4)

Smithrim (1998) explains that she was interested in 'changes in the role of singing over time' and in order to study this, she recruited 'girls and women who have lived in the same small city and attended the same church for most of their lives':

'Because I was interested in singing in women's everyday lives, I did not seek out women who would identify themselves as performing singers but neither did I reject the two who do happen to call themselves singers. I interviewed each girl and woman in their own homes. Interviews included questions about earliest musical memories, song of childhood and adolescence, the songs of both wars, the lullabies; and how they feel about singing and their own singing voices. I also asked about the place of singing and songs in their spiritual life, their emotional life, their social life and their family life.' (1998:218)

Patteson (2000) writes as a singing teacher about the experiences of women's taking lessons with her. As a background to the study, she explains that many of the women she worked with found that singing lessons had effects which:

'...reached far beyond the weekly session we spend together and beyond the growth of their singing voices, having what they have described as transformative effects on their whole lives.'  
(2000:184).

The aim of her research therefore, was:

'...to understand how, for eight women ranging in age from thirty-three to fifty-five years, taking singing lessons has transformed how they feel about themselves and how they live their lives. I chose as my focus the singing lessons, rather than the specific activity of singing, because my study participants told me that what happened

between them and me in the pedagogical setting of the lessons was essential to their processes of change.’ (2000:184)

The breadth and depth and complexity of the issues addressed by these authors is apparent from these extracts – as is the very particular features that each was interested in, and strikingly, the way in which singing as an activity is placed within much wider contexts of history, culture, personal biography and growth and the quality of relationships between women. The papers are concerned with singing, but for Joyce the focus is on the role of singing in ‘some process of change or growth’ in women’s lives; for Smithrim it is on the role of singing in everyday life among women who would not identify themselves as ‘singers’, and for Patteson, it is the ‘singing lesson, rather than the specific activity of singing.’ These studies serve first and foremost, therefore, to highlight the multiplicity of embedded contexts and issues raised by any attempt to think about the value of singing in relation to personal wellbeing and health.

Without any intended irony, the best that can be done with these sources, in the context of this review, is to offer some extracts in which the authors highlight benefits for wellbeing and health.

Joyce (1996) sums up well the sense that singing is an activity which engages the body and mind in a holistic way as well as connecting us with other people:

‘In many dramatic ways, singing takes us into our bodies and puts us in touch with our emotions. Singing can relax our body, release physical tension, and provide a vehicle for safe emotional release. It generates physical energy and aerates our bodies through deep and sustained breathing. All our physical senses are stimulated by the process of singing which in turn produces heightened states of sensitivity and arousal. (...) Singing can connect us internally to our own complete inventory of learning capabilities. But even greater possibilities exist when a group of individuals sing en masse. While exercising our individual capabilities, particularly the capacity for relational learning, all members of a group contribute to building a synergistic spiralling of collective energy, insight, and creativity which is extremely compelling and effective.’ (1996:256)

We have here, in effect, a basic theoretical model for understanding the ways in which singing can potentially impact on personal wellbeing and health.

Smithrim (1998) provides a remarkable life-span view of singing, highlighting women’s experiences of singing and songs from the earliest years through to old age. Some of the participants felt that their singing voices ‘were not very good’, but all of them placed a high value on singing in their lives. The following quotation sums up powerfully the significance of singing for the women in her study:

'... many spoke passionately about how singing affects their lives. Singing, for these women, was a form of self expression, a stress release, a transcending or transforming activity, a confidence booster, a powerful binder with other women, a support, a sedative and a comfort and joy to others.' (1998:125)

Patteson (2000) states that all of the women in her study identified a connection between their 'physical voice' and their 'metaphorical voice'. Developing their singing voice, enabled them to take steps towards 'overcoming silence and reclaiming power' in their lives. In this sense, taking singing lessons was transformative for all of the women in the study and helped them in addressing serious challenges to their sense of mental, emotional and social wellbeing. Patteson sums up some of these challenges as follows:

'Brenda, Sarah, and Evelyn were able to begin or to continue the process of recovery from the effects of childhood sexual abuse. Kathy said that the singing lessons provided her with an experience of joy and feelings of renewed hope in the midst of clinical depression. Fiona found an ability to grieve the death of her husband more thoroughly. (2000:198)'

#### **4.1.2 Three more recent qualitative studies**

##### **Ashley (2002)**

Ashley (2002) reports findings from a study of choir boys (aged 10-14) singing in 'a major city centre parish church' in the UK. This is the only study found in the search which addressed issues of singing and wellbeing in relation to children.

The aim of the study was:

'... to gain an understanding of boys' experiences of singing in a church choir and what it meant to them, and to explore the implications of these experiences to their sense of identity and personal wellbeing.' (2002:182)

The study considers influences operating on boys' gender identity formation in relation to music, and Ashley argues that:

'... a major source of proscription against boys' participation in music, and singing in particular, would appear to be the culture of boys themselves. Hegemonic or dominant forms of masculinity (Connell, 1995) seek power and cultural supremacy through a social construction of masculinity as performative and relational (Frosh et al., 2001). This means that performances such as football skill, considered in popular boys' culture to be "masculine", receive approbation, whilst performances such as singing, considered "feminine", are generally proscribed.' (2002:181)

Ashley describes the research in the following terms:

'An ethnographic, case study approach was adopted, and information gathered through participatory observation, diary keeping and interviews. The diary recounted an "insider's view" (Beckerlegge, 1998) of the day-to-day incidents which made up the life of the choir. Each boy was interviewed in depth at home on two occasions, and further interviews were also conducted with the same boys in groups of three.' (2002:182)

From a methodological point of view, Ashley's aim was 'to treat the boys, not as "objects" of research, but as co-participants in its process.' Each boy was interviewed in depth at home twice and group discussions were conducted with the boys in groups of three. The account of data analysis is somewhat weak however. Ashley simply states that 'a number of key themes emerged from the data gathered', and no further information is given on the method employed, nor checks on reliability and validity of interpretation. In the method section it is claimed that 'respondent validation' played 'a key role' in the research, but this is not elaborated.

Principal findings on the boys' involvement in music and links to well-being were that the boys showed a deep appreciation of and engagement with music, and exhibited many features of high personal wellbeing, including 'the social competence to circumscribe the proscriptions of "macho" culture.' (2002:180)

Ashley provides a number of quotations from the boys, which show how the boys experienced an engagement with the music they were singing, and how this was often explicitly linked by them with an experience of wellbeing:

*'The Durufle Requiem was boring at first but it grew on me. Now I've had it on my brain all week.'*

*'It makes me feel happy, Howell's Here is the Little Door make me feel different.'*

*'Yes, it's a feeling you can't put into words.'*

*'It makes me feel better, especially if it's Greater Love by Ireland – it moves me to tears, because I like it so much.'* (2002:183)

### **Bailey and Davidson (2002a, 2005)**

The papers by Bailey and Davidson (2002a, 2005) which report qualitative research undertaken by Bailey for her PhD are among the most important contributions to the literature on singing and health. Their significance lies not simply in the fascinating data gathered through interviews and discussion, but also in their theoretical sophistication, drawing both upon the Flow Theory perspective developed by Csikszentmihalyi (1997) and elaborating their own

specifically focused 'theory of the positive effects of participation in group singing and performance.' (2005:298).

Bailey and Davidson (2002a) offer a full discussion of Csikszentmihalyi's perspective, but the essential meaning of 'flow' is explained succinctly as follows:

'According to Csikszentmihalyi meaningful life demands focussed attention and when attention is directed on a sufficiently complex task we experience total mental absorption or "flow": "Because of the total demand on psychic energy, a person in flow is completely focused. There is no space in consciousness for distracting thoughts, irrelevant feelings. Self-consciousness disappears, yet one feels stronger than usual." (p.31)' (2002a:225)

A second element in Csikszentmihalyi's model is the importance of social relationships and interaction in the promotion of mental wellbeing:

'Csikszentmihalyi (1997) proposes that social isolation encourages distracting and distorted thoughts and contributes to the development of emotional disorders. His research indicates that people experience more negative feelings when they are alone, and that negative feelings dissipate in the company of others. (...) Whereas loneliness facilitates psychic entropy, social encounters (...) require orderly and directed attention that is similar to, though not as intense as, experiencing flow.' (2002a:225-6)

Bailey and Davidson apply this perspective in an illuminating and convincing way to the experience of group singing:

'In group singing there is a continual feedback concerning note, rhythm and pitch accuracy, proper pronunciation of words and the blend of various sections. Each rendition of a song conveys new information which may be applied in subsequent repetitions resulting in a sense of progress, achievement, and an increase in self-esteem. In addition to the cognitive component, group singing also provides participants with opportunities for social engagement. As suggested in flow theory, social engagement provides opportunities to direct attention away from the self. (...) Therefore the effects of active participation in a musical activity, such as group singing, may provide opportunities to experience the positive effects of mental stimulation and social engagement as described in flow theory.' (2002a:226)

Bailey's initial study investigated the experiences of homeless men participating in a choir established in a shelter in Montreal in 1996. From small beginnings the choir stabilised to having a membership of approximately 20, and went on to achieve considerable success, performing in many concerts both nationally and internationally. Bailey and Davidson argue convincingly that given the backgrounds of the men involved in the choir, and

the successes they had achieved, that interviews with members of the choir could provide valuable insights 'relating to the effects of the group singing experience' and in particular whether it provided 'some psychological or physiological advantage.'

Participants in their study were 7 of the 17 active members of the choir (average age 52 years). The participants were all native French speakers, but they were interviewed in English. The interview itself was wide-ranging asking participants about their early family life, education and employment, and how they became homeless, before moving onto their experiences within the choir and the effects they felt it had on them.

A central theme emerging from the interviews was that men experienced therapeutic benefits from the experience of singing in the choir. It was also clear that the factors of mental stimulation and social engagement were important factors for them in giving rise to these benefits. A selection of relevant quotations from the men themselves, serves to highlight these issues:

*'When we perform, just in the subway sometimes, I don't feel like going, but after I go I feel so good you know, it refreshes you. I would practice eight hours a day I like it so much, so when there's no practice I don't feel good... I'm not the only one, we need it now, it's nuts you know, we need it. It is a therapy I think myself. '(Henri) (2002a:236)*

*'For me it's very therapeutic as before coming here I was in very bad shape emotionally. (I have found) a way out of my traumas. Because if you always live in the past it's not good. Some of us are so uncomfortable, we talk about this one and that one and we never talk about the real thing, you know. These days, I suffer from arthritis in the knees, but the minute the music starts I don't feel my arthritis anymore. The emotional part of it will last eight to twenty hours, and that's worth a lot of money. For me it is a drug.'* (Raoul)(2002a:237)

*'Physically, mentally it's different as day and night (comparing singing to not singing), if I don't sing during the day I feel I'm depressed, but if I sing there's something magic in me, I do what I love to do and that is magic. I'll die if I don't sing, I would go depressed if I don't sing. Usually we sing at night. I'm so high because of having singing. I'm on a trip, it's a drug for me, it's a real drug, a natural drug.'* (Louis)(2002a:238)

Bailey and Davidson believe that the homeless men gained 'emotional, social and cognitive outcomes' from group singing, and sum up their findings in terms of four themes which emerged repeatedly in the men's accounts:

- Group singing alleviated depression and enhanced emotional and physical well being

- Performing to an audience encouraged a sense of personal worth and provided a means of re-engaging with wider social networks
- The choir provided a supportive context for the men in which they could develop their social skills and achieve collective goals
- Singing is mentally demanding, and required the men to concentrate and learn new material in order to perform. Such concentration also directed their attention away from internal preoccupation with their problems

Bailey and Davidson employ these four themes in the construction of 'a working model of a theory of the adaptive characteristics of group singing' and suggest that their model 'may provide a useful preliminary framework to guide research in an area of music that generally has been neglected.' (2002a:246) This study provides a great deal of insight into the potential power of singing in relation to wellbeing and health, but despite its many merits, it has a limited focus on members of one very particular, and perhaps unique choir, and the numbers of participants in the study was also very small (7) and restricted to men.

To address these limitations, Bailey and Davidson (2005) build on their initial research in a second study of members of a mixed choir (7) established in a disadvantaged community setting, and a third investigation of middle class singers involved in more traditional classical choral singing (8). Qualitative data was again collected and subject to thematic analysis. Findings from these studies are considered in considerable detail in relation to the first study of homeless men, and their analysis leads to a revision of the 'working model' presented in the earlier paper.

The four major themes identified in the 2002a paper are considered to be equally applicable to singers in all the three studies, irrespective of social context and the character of the repertoire being sung (i.e. Clinical type benefits, Group Process, Choir/audience reciprocity and Cognitive stimulation). Both disadvantaged and more privileged singers highlighted the broadly 'therapeutic' value of participation in singing, particularly in relation to creating energy, positive emotional experience and relaxation. For other themes, in contrast, some important differences emerged, particularly in relation to the cognitive dimensions and impact of singing in a group. For the more marginalised participants, singing provided a stimulating activity which helped to promote concentration and an ordering of their inner mental space. For the middle class singers, in contrast, a greater stress was placed on developing musical knowledge and skill which enabled them to meet the challenges of classical repertoire and gain a sense of achievement.

Bailey and Davidson conclude:

'This research is a first step in understanding the importance of participatory singing experiences for amateurs at many different levels of musical training and ability. It illustrates that singing and



performance can produce satisfying and therapeutic sensations even when the sound produced by the vocal instrument is of mediocre quality. In fact, there may be very little difference in the enjoyment of generating musical sounds at the most professional and most amateur levels.' (2005:299)

### **Silber (2005)**

Silber reports a study focused on the potential benefits of a group singing activity in the context of a women's prison in Israel. In some respects, the study is similar in its aims and scope and to that undertaken by Bailey and Davidson, particularly in relation the 'therapeutic benefits' that could come from group singing. It differs from the Bailey and Davidson study, however, in that Silber led the singing group in the prison. She was thus in a position to document the processes involved in working with the women and identify the ways the palpable benefits of singing experienced by the participants, were brought about:

'The goal of the study was to elucidate the points of interface between the ordinary workings of a multi-vocal choir, and the therapeutic needs of prison inmates. The study design was qualitative, the conclusions being based on the observed effects of the project on its subject participants, analysed in consultation with a clinical psychologist to provide further perspective and insight.'  
(2005: 255-6)

The Silber study is unique within the corpus of research papers identified for its detailed attention to the workings of the choir from a social and musical point of view. The choir itself was 'a multi-part vocal ensemble' which met weekly over a period of eight months. Each rehearsal lasted 90 minutes and sessions were devoted to learning songs and 'developing the participants' choral skills.' Three performances were also arranged to serve as goals and 'milestones in the group process.'

A range of methods of documenting the work were employed by Silber in order to gain a sense of perspective on the process and reach 'objective conclusions.' Methods included the videotaping of sessions, keeping a journal, observations of relationships within the choir and discussion and interviews with participants. In analysing all of this material, Silber identifies three aspects of the choir experience which she believes provided some therapeutic benefit to participants:

- The vertical relationship between conductor and choir members
- The horizontal relationship between choir members
- The individual's role as choir member

Each of these three aspects is discussed in some detail, and their role in mediating therapeutic change illustrated. What is very striking about the

discussion is the way in which Silber links specific musical concepts, practices and skills, to the encouragement of positive changes in the behaviours and attitudes of the women participants.

With respect to the first aspect, the vertical or leadership relationship between herself and the participants, Silber highlights, the power of her hand signals as a choral director, in context of the development of the choir. This illustration is so striking, it is given in full:

‘The most revealing finding along the horizontal axis was the added significance of the very didactic and functional choral gestures and hand signals whose application in an ordinary choir are limited to the musical realm. In the present choir, these signals became a fascinating medium for communication of authority. As choir members became comfortable with conducting language, I increasingly employed musical gestures not only to convey choral instructions, but to distinguish acceptable behaviour from unacceptable. One incident in particular stands out as being representative of this usage and its potential applications. When one of the inmates did not respond to my request to be silent, I asked her to demonstrate how I conduct a ‘cut-off’, or a musical stop. A different inmate responded, correctly, by closing her hand, and everyone followed suit. This served to defuse the head-off between the woman who had made the disturbance and myself, and refocus attention from the *ad hominem* reprimand to the functional. The group then agreed that whenever this sign was given in future, we would all be silent.’ (2005:259)

With respect to horizontal relationships, the challenge facing Silber was to create a cohesive group. One of the simple, but highly effective musical techniques Silber employed to help achieve this was the use of ‘call and response’:

‘Knowing that common routine that is ritualized can have a uniting effect, I introduced a bouncy three-word song, in two-part harmony, which became our ‘anchoring’ piece that we sang at the beginning and end of our sessions. Significantly, the song was entitled ‘*Chazak*’ (‘Be Strong’), a theme, I later discovered, that was extremely important to the choir members, given their experience of the world as a place in which only the strong survive. Paradoxically, the call-and-response format of the song also elicited a more tender side of the singers. Whenever the choir sang ‘*Chazak*’ – where one voice would begin and the other would follow in a harmonic response – a kind of musical dialogue would automatically occur, resulting in an almost palpable easing of the tensions that had accumulated in the course of the prior week and a heightened sense of unity and trust among the singers.’ (2005:260)

A focus on breathing and head voice, provide further striking examples of how specifically vocal /musical processes could function 'horizontally' within the group of women to address aggressive feelings and behaviours. The accounts offered by Silber show very concretely, the intrinsic interplay between the bodily mechanisms which power the voice, and the qualities of the sound produced – and processes operating on a psychological and social level within and between individuals:

'First, I worked with the choir members on the activation of the diaphragm as a means of achieving the steady flow of air necessary for proper singing. This practice of breath control also served as a potentially useful method for releasing tension, since it slows the heart rate and calms the nervous system, resulting in a more relaxed state.'

'The second device that I employed to deal with certain choir members' belligerent behaviour was the introduction of the higher voice register, which is a gentler and softer head-tone quality. In their music, as in their prison experience – and for many of the inmates, in their lives before prison as well – they were unaccustomed to this understated and non-aggressive voice. They did not sing in that voice, nor was it common in the popular music they enjoyed. (...) At first, the choir members (...) objected to singing in this delicate manner, fearing that audiences would think they had no 'power' in their voices. (...) I reasoned with them in keeping with their own logic, arguing that their singing in the higher register was actually perceived as 'stronger' precisely because it was more pleasant and unforced. Ultimately this message was reinforced in performances when the audience responded enthusiastically to this voice. We decided to refer to the higher register sound as 'the new voice.'

On an individual level too, the experiences of group singing appeared to be beneficial in a variety of ways. Group singing gave women a great sense of self-control in relation to their contributions in achieving a group product; it helped to encourage a sense of delayed gratification in working toward longer-term goals of performance, it helped women develop sense of 'empowerment' and 'ownership' and above all promoted a sense of self-esteem and self-worth.

Silber's study is clearly small scale and has a particular focus on female prisoners, but without doubt it highlights important issues of relevance any discussion of singing, wellbeing and health – not least the need to understand how the intrinsic features of group singing as an activity might impact on an individual's sense of wellbeing and positive health. More than any other study in the corpus of research considered in this review, Silber succeeds in highlighting the specifically 'musical' processes through which the links between singing and health can be made.

## **4.2 Surveys of singers using specially constructed questionnaires**

A number of studies have undertaken surveys of choral singers using specially constructed questionnaires with a focus on wellbeing and health (Hills and Argyle, 1998a,b; Beck, Cesario, Yousefi and Enamoto, 2000; Clift and Hancox, 2001; Bailey and Davidson, 2003a), or related relevant concepts such as quality of life (Hillman, 2002) and social capital (Louhivuori, Salminen and Lebaka, 2005).

As with the qualitative studies just considered, the studies are diverse in terms of objectives, samples and the specific details of the instruments employed to gather data. Consequently, a synthesis of findings is difficult. The studies will be briefly described in order of publication, and some of the common themes emerging will be highlighted. Reference will also be made to the studies reviewed in the last section where appropriate.

A further study to be considered in this section by Sandgren (2002) is somewhat different in being concerned with professional opera singers, and having a focus on problems with health associated with singing, rather than benefits.

### **4.2.1 Studies of choral singers**

#### **Hills and Argyle (1998a, 1998b)**

Hills and Argyle (1998a, 1998b) report two studies in which specially devised scales were used to assess emotional experiences associated with leisure activities. In the first study, comparisons were made between 71 adults who were members of musical groups (primarily choirs), 121 who attending church services, 54 who engaged in both activities and 68 who belonged to neither. Two scales were employed. The 'religious experiences scale' was made up of 25 6-point items designed to measure aspects of 'relationships with others', 'positive mood' and 'transcendental experience.' The 'music scale' was similarly constructed with 24 items, including specific statements concerned with 'musical satisfactions and relations with others.' Eleven of the items were common between the two scales, to allow for comparison. For the group who were members of both musical groups and churches, comparisons showed that feelings were significantly more intense in response to music than church service attendance on four items – 'excitement', 'feeling uplifted', 'joy/elation' and 'loss of sense of self.'

The concerns of this study to compare groups is of limited relevance in the context of this review, but what is of interest, is the result of a factor analysis of the music scale. Table 8 reports the five factor solution found for the music items, with the percentage variance explained by each factor. Factors were extracted initially using principal components analysis, followed by Varimax rotation for factors with eigen values greater than one. Unfortunately factor loadings are not reported, and so the degree to which simple structure was obtained cannot be assessed. Nevertheless, the interpretation of the factors appears reasonable given the content of the items associated with each of

them. Of particular interest is the fact that the first factor is associated with 'wellbeing' and has loadings from items concerned with 'mental' and 'bodily' wellbeing.

Table 8 Factor analysis of the music scale items (Hills and Argyle, 1998a)	
Factor (shared variance)	Items
Factor 1: Wellbeing (33.6%)	Being bathed in warmth and light Bodily wellbeing Feeling uplifted Excitement Mental wellbeing Achievement Joy/elation
Factor 2: Mystical experience (9.6%)	Timelessness Glimpsing another world Self-discipline Loss of sense of self Identification with performers
Factor 3: Social (8.0%)	Enjoying company of others present Getting the best out of one's self Challenge Taking part in a shared performance Positive feelings about life
Factor 4: Entertainment (7.0%)	Reminder of happy occasions Entertainment Relaxation/calmness Recognising the familiar
Factor 5: Intellectual / musical (5.4%)	Appreciating a good performance Pleasure in musical structures Mental stimulation

In their second study, Hills and Argyle (1998b) compare the 'positive moods' generated by four common leisure activities in a sample of 275 adults. Among these, 111 were involved in sport and exercise, 76 in music groups (primarily choirs), 127 in church services and 119 in watching television soaps. As the subgroup numbers add up to more than 275, participants were clearly involved in more than one activity. As with the 1998a study, the concerns of the authors were primarily to compare the experiences of different leisure activities is of limited relevance to the current review. Of more interest, however, is that again the scales are subject to factor analysis. In the later study, the music scale is briefer and consists of only 8 as opposed to 24 items in the initial study. Consequently, the factor solution consists of two factors as reported in Table 9.

Hills and Argyle (1998b) neglect to report the variance associated with each of these factors and again fail to report item loadings. Nevertheless, the two analyses are consistent in revealing two factors associated with the experience of participation in music groups (primarily amateur choirs). The first and strongest factor in both analyses concerns individual wellbeing, and the second is concerned with the social character and benefits of music making as a group activity. However, it is not clear whether the samples in the

two studies are different or whether there is some overlap (in 1998a N=71 for the music group and 1998b N=76 for the music group). If the latter, it would not be surprising to find similar factors emerging.

Table 9 Factor analysis of the music scale items (Hills and Argyle, 1998b)	
	Items
Factor 1	Feeling uplifted Joy/elation Excitement Bodily wellbeing
Factor 2	Enjoying the company of others present Taking part in a shared performance Timelessness Entertainment

### **Beck, Cesario, Yousefi and Enamoto (2000)**

The study by Beck et al. (2000) is primarily significant for its assessment of changes in immunoglobulin A and cortisol levels in response to singing. In addition, however, they gathered considerable self-report data using a specially constructed questionnaire and specific rating scales.

The participants in the study were members of a large professional choir, the Pacific Chorale, with 180 members. Singers considered ineligible for the study were excluded and from the remaining 152 singers, 41 were randomly selected to participate, 23 females and 18 males with a mean age of 46 years, and average experience of singing in choirs of 37 years. No justification is given of the sample size in terms of power.

The first questionnaire is described as a 'Singers Emotional Experiences Scale' which consists of 28 items with five-point response scales, designed to assess 'typical experiences connected with rehearsals or performances, before, during and after singing.' In addition, choristers completed a 10-item 'self-report form' for two specific rehearsals and one performance 'to measure personal perceptions and evaluations of experiences associated with singing on *particular days*.' (2000:95)

The 'Singers Emotional Experiences Scale' revealed clearly that singing is associated with strong feelings, with singers tending to agree strongly with positively worded statements and disagree with negatively worded statements. More than half of the singers agreed or strongly agreed with 13 of 19 positively worded items, and disagreed or strongly disagreed with 3 of 9 negatively word items (see Table 10).

Interestingly, two of the items in Table 10 refer to 'personal well-being' and 'health', with more respondents agreeing with the former (68.3%) than with the

latter (56.1%). Some members of the choir clearly believed that singing benefits personal well-being, but stopped short of claiming benefits for health. The self-report form data served to assess such feelings as stress, excitement, anxiety and enjoyment associated with two rehearsals and one performance. The results revealed that ‘performances were associated with both positive and negative emotions more than were rehearsals’ (2000:100), i.e. the performance elicited strong levels of enjoyment and excitement, but also (not surprisingly) more anxiety.

Table 10: Responses to ‘Singers Emotional Experience Scale’ items  
(Beck et al. 2000)

Positive statements	% agreement
Singing is very important to me	92.5
Singing has made me a happier person	80.5
I feel exhilarated after a performance	70.7
Singing has contributed to my personal well-being	68.3
Singing is usually a release	65.9
Singing gives me an identity as an artist	65.9
Singing gives me a kind of high	63.4
I feel something deep inside me when singing	58.5
I feel strong emotions when singing	56.1
I feel singing has made me healthier	56.1
Singing gives me a strong spiritual feeling	51.2
After singing I usually feel energized	51.2
Negative statements	% disagreement
Singing generally lowers my mood	57.5
I don't usually feel strong emotion when singing	56.1
I don't usually get a spiritual feeling when singing	53.7

The Beck et al. study is primarily focused on the singers’ experiences in the context of rehearsals and performance, but they did also gather accounts which indicate that some singers saw singing as making a significant contribution to the general quality of their lives. Beck et al. do not report on these findings fully, but they do give some of the comments made by respondents with respect to coping with general life stresses:

‘An analysis of qualitative responses to open-ended questions on the SEES revealed that many singers (approximately 25%) specifically viewed the benefits of singing as highly useful in coping with stress (...) For example, one singer said: “When I’m stressed all I have to do is start singing and the tension is gone.” Another singer said [During singing] “I’ve cried out personal loss and celebrated joys.” Another singer wrote “It is for me very healing. I can come into a rehearsal or performance ‘world weary’ and leave refreshed and renewed.” And finally one singer wrote: “It gives me something to focus my energy on that takes me away from the worries of the day job.” Thus, for many singers it is likely that the stress-reduction powers of singing are considerable.’ (2000:104-5)

## Clift and Hancox (2001)

Clift and Hancox (2001) report two studies investigating the perceived benefits associated with choral singing. The first study was qualitative and exploratory in character and involved asking choral singers to write answers to simple open questions about their experience of singing:

‘Do you feel you have benefited personally in the following ways (four areas were specified - physically, emotionally, socially, spiritually) from being involved in this choir so far? If ‘yes’ please explain how.

Are there any ways in which you think that participating in this choir could be ‘good for your health’? (2001:250)

Participants in the study were 84 members of a University College Choral Society, and a majority of respondents agreed that singing gave them social (87%), emotional (75%) and physical benefits (58%). Just under half also agreed that singing gave them spiritual benefits (49%).

Of greater interest, however, are the specific ways in which choristers believed they benefited from singing in a choir, and Table 11 (based on Table 2 in Clift and Hancox) presents the ten most commonly expressed themes.

Table 11: Perceived benefits of choral singing – themes identified in the first Clift and Hancox (2001) survey		
	n	%
Social benefits (n=72)		
Have got to meet/know more people through singing	46	64
Have made new friends	13	18
Emotional benefits (n=59)		
Makes me feel positive/good/happier/raises mood	36	61
Emotional workout/induces emotions	7	12
Helps to release stress/reduce stress/lessen tension	7	12
Physical benefits (n=45)		
Increased control over breathing/improves breathing	20	44
Wakes me up/feel more alert/energised/active	9	20
Improved posture/improved stance	8	18
Exercises lungs/improved lung capacity	6	13
Spiritual benefits (n=37)		
Spiritually up lifting	10	27



Clearly, the three most important issues identified are increased social connectedness, increased positive mood state and improved breathing. Responses to the second, more specifically focused question on benefits for health reinforce the main themes indicated in Table 11. A total of 74 respondents provided answers to this question, and 41% mentioned 'improved lung function and breathing', 30% referred to 'improved mood and feelings of happiness' and 22% suggested that singing can help to 'reduce stress'.

Interestingly, some of the comments on the questionnaire pointed towards a holistic concept of health in which the physical, psychological, social and spiritual components are inter-related and mutually reinforcing:

*"It's good for one's soul and what's good for your soul is good for your body."*

*"If the physical side is related to your spiritual side, then it can do only good. Healthy mind, health body etc."*

*"To experience emotional well-being is (must be!) very good for health. Music is a natural thing for humans." (2001:252)*

On the basis of this initial exploratory survey, Clift and Hancox constructed a 32 item questionnaire. The wording of items in the questionnaire was closely guided by the statements made by respondents in the first survey. The structured questionnaire was then completed by 91 members of the same choral society. Over 40% of respondents strongly agreed that 'singing helps to make my mood more positive', 'singing is a moving experience for me sometimes', 'singing makes me feel a lot happier' and 'singing is good for my soul' thus emphasising the significance of singing primarily on an individual emotional level for participants.

The questionnaire was then subject to principal components analysis followed by an Oblimin rotation which allows factors to be correlated. Six dimensions of benefit emerged associated with choral singing. These were labelled (in order) as benefits for: well-being and relaxation, benefits for breathing and posture, social benefits, spiritual benefits, emotional benefits and benefits for the heart and immune system. The first three factors can be taken to reflect the three components of wellbeing identified in the World Health Organisation definition of health – mental wellbeing, physical wellbeing and social wellbeing. The holistic and interactive positive view of health expressed in the WHO definition is reflected in the fact that these separate components are correlated with one another. There is also a clear link between the set of components emerging from this analysis and the model of positive benefits of group singing emerging from the work of Bailey and Davidson (2002a, 2005). It is instructive to look at the first and largest factor (wellbeing and relaxation), in more detail. Table 12 (based on Table 5 in Clift and Hancox) gives the items with loadings above 0.6 on this first factor.

The initial items highlight the immediate impact of singing on mood and sense of happiness in the setting of a rehearsal or performance, but subsequent items point to a wider context and impact – not only in helping to release stresses and forget worries from outside the singing session itself, but also in reinforcing a more general sense of wellbeing and positive view of life which goes beyond the activity of singing. This initial component appears to tie in well with the work of Csikszentmihalyi (1997) highlighted by Bailey and Davidson, and might even be described as reflecting elements of an experience of ‘flow’ in the context of choral singing.

Table 12: The first component emerging from the second Clift and Hancox (2001) survey

Items	Loading
Singing helps to make my mood more positive	.88
Singing is something I look forward to	.80
Singing helps me to relax	.78
Singing helps to improve my mental wellbeing	.77
Singing makes me feel happier in myself	.76
Singing helps to make feel calmer	.70
Singing reduces stress	.67
Singing helps me to feel more awake and alert	.67
Singing helps to give me a positive attitude towards life	.63
Singing helps me to forget worries	.62

### Hillman (2002)

The participants in this study were members of a large community choir (Call That Singing, CTS) established in Glasgow in 1989 as part of its European Capital of Culture celebrations that year. The choir was open to people over the age of retirement and proved to be so popular and successful that it continued to meet and was still functioning in 2001 at the time of this study. Hillman explains that her research arose ‘through observation of people involved in CTS and their articulation of the perceived positive effects on their general health and wellbeing’ and the study was undertaken ‘to determine whether there is any evidence to support the general view that “singing is good for you.”’

A detailed questionnaire was sent to 100 members of the choir and a total of 75 usable questionnaires were returned (60 females, 15 males). Included within the questionnaire were asked to make judgements of their health and wellbeing before and after joining the CTS choir. The focus of the study was therefore on perceived changes over time. Participants had been members of the choir for between six months and eleven years, but this factor was not taken into account in analysing the results. A further limitation is clearly the possibility of demand characteristics which might lead participants to understate their health and wellbeing before joining and overstate it at the time of completing the questionnaire.

Nevertheless, Hillman's findings are interesting in showing a variable pattern of perceived change across the various scales used to assess wellbeing and health. Ratings of general 'quality of life' showed the most marked change with 76% giving a rating of excellent/good 'before' joining and 94% giving the same rating 'after' (chi-square, 11.34,  $p < 0.001$ ). Ratings of emotional wellbeing also improved (70% to 90% excellent or good,  $p < 0.05$ ), but ratings of physical health showed a non-significant decline (78% to 71% excellent or good).

Of great interest than the structured data, however, are the comments given by participants in response to an open question on 'marked changes' experienced as a result of being part of the choir. These included comments on beneficial changes experienced in wellbeing and health:

*'Being part of CTS helps to make me forget all my silly aches and pains when I need to the most. I would miss it very much, if I had to give it up. Singing together is very therapeutic for me (married woman)'*

*'I was alone and withdrawn. I had isolated myself. Through my involvement with CTS my self-esteem has improved. I love singing. I am surrounded by like-minded people and have learned to live in the present (widow).'*

*'At a time in my life when I was very low due to my husband's death CTS filled a very big gap (widow).'*

*'CTS is a kind of therapy for most people who attend, especially people on their own, to meet friends and other people (married woman) (2002:166-7)*

It is striking that all of the comments quoted by Hillman come from women, and not surprisingly, given the age of the sample and greater life expectancy among women, many of the women are widowed.

### **Bailey and Davidson (2003a)**

Bailey and Davidson (2003a) aimed to explore the extent to which different levels of participation in music are perceived to carry health benefits in a holistic sense. In an admirable progressive spirit, they build upon their own work with members of homeless and disadvantaged choirs, and studies by Beck et al. (2000), Kuhn (2002) and Kreutz et al. (2004) and design an investigation to compare the perceived benefits of 'active' participation in music (singing in a choir), 'passive' listening to music with others, and 'passive' listening to music alone.

Choristers from three choirs ( $n=121$ ) participated in the survey and were invited to complete a 100-item questionnaire described as measuring attitudes related to music. Included in the questionnaire were 3 sets of 22 items concerned with the 'holistic health effects of music in the three participation

categories.’ Additional items relating to choir practices, performances, voice quality and conducting techniques were included to ‘camouflage’ the purpose of the questionnaire. The study is similar to that undertaken by Clift and Hancox (2001), but developed further by asking the same group about the perceived benefits of listening to music, in addition to performing music as singers in a choir. It also corresponds to some extent to the approach adopted by Hills and Argyle (1998a,b) in assessing the feelings associated with different leisure activities. Bailey and Davidson give the following illustration of how the items were linked to the three levels of participation:

‘(1) Singing in a choir usually improves my mood, (2) Listening to music alone usually improves my mood, and (3) Listening to music with others usually improves my mood.’ (2003a:221)

Items were written related to emotional, physical, cognitive, social benefits and life satisfaction. For each set of items, half were worded positively and half negatively. Items in each set were treated as separate scales and alpha coefficients were consistently high (close to 0.9 in each case). Bailey and Davidson then compared the three sets of items in terms of the levels of agreement and disagreement by survey respondents. Their central finding is summed up as follows:

‘The differences in the ratings of the items of the 3 music involvement scales suggest that active participation in choral singing was considered a more holistically beneficial activity than listening to music alone or listening to music in the company of others. Also, segregated listening was perceived to be more holistically beneficial than social listening.’ (2002a:222).

Bailey and Davidson go on to suggest that many of the group singing items that received higher ratings than equivalent listening items, indicate that singing ‘promotes heightened arousal on a variety of behavioural dimensions.’. These items were:

- Improves mood
- Is an exhilarating activity
- Gives me a sense of achievement
- Is a creative experience
- Gives me a kind of high

By contrast, the listening alone items which received the highest ratings were ‘indicative of processes which promote stress reduction and restoration of a homeostatic state.’ These items were:

- Reduces stress
- Releases suppressed emotions
- Is physically relaxing
- Releases tension

Bailey and Davidson conclude from their findings 'that different levels of music participation can have differential yet beneficial effects.' There is thus a need for further research to develop 'more effective uses of diverse types of music activities to promote holistically healthy behaviours and practices which are appropriate to the requirement of therapeutic, industrial and educational environments.' (2003a:223)

### **Louhivuori, Salminen and Kebaka (2005)**

The study by Louhivuori et al. is unique in the corpus considered in adopting a cross-cultural approach to the experience of choral singing. Data were gathered from members of choirs in Finland and South Africa by questionnaire with a focus on the social dimensions and value of the choral experience, guided by the concept of 'social capital' as developed by Putnam (2000). The aim of the study was:

'... to find out how the social capital of participants develop through musical group activities and what the meaning of music hobbies is in the development and growth of trust between the members of the community.' (2005:81)

Louhivuori et al. provide limited information about their methods and the participants involved and say only that 'over 500 choir singers' were involved with a wide age and educational range, and that 'all the usual types of choirs are represented: youth/male/female choirs, mixed choirs, church choirs and senior choirs.' (2005:83) The presentation of findings is also somewhat weak, with data reported visually rather than in tabular form, and no detailed information given on the statistical analysis employed.

The major part of the study is focused on respondents' motivations for being part of a choir, and benefits they experienced, primarily in relation to music (e.g. increased knowledge and skills). However, a further question asked about the importance of personal relationships within the choir, and this issue is likely to have a bearing on an individual's sense of social well-being and positive connectedness with others. Here an interesting cross-cultural difference emerged, with 75% of South African singers rating personal relationships as 'very important', compared with 40% of Finnish choristers. A further question is of more direct relevance to wellbeing and health asked participants whether they felt able to share or discuss personal matters (including physical and mental health) with at least one member of their choir. The data gathered for this question are presented as a figure in the paper (Figure 7) and precise percentages are not reported, but the pattern appears to be that Finnish singers are more willing than South African singers to discuss physical health matters (approximately, 85% to 70%). In respect of 'mental health matters', in contrast, Finnish singers appear less willing to discuss such issues than South African singers (approximately 60% to 45%). The difference in levels of willingness to discuss physical and mental health is thus particularly marked among the Finnish singers (85% physical health, 45% mental health) and may reflect a greater sense of stigma associated with mental health.

Louhivouri et al. give the following summing up of the results emerging from their study, which interestingly highlights the satisfaction of multiple human needs as a basis for the enduring attraction of choral singing:

'In general the results give an impression that singing in a choir supports both individual and communal needs of people. At the same time strong artistic experiences can be felt and social needs are fulfilled. The ability of choir singing to satisfy many and different kinds of needs of human beings might be one of the main explanations of the popularity of choir singing (2005:92)

#### **4.2.2 A study of professional opera singers**

##### **Sandgren (2002)**

This study is unique in its focus on professional opera singers and in exploring the psychological and physical demands associated with pursuing a highly demanding career as a singer. In this respect, the study has less relevance to the broad concerns of this review on the health benefits of group or choral singing, but nevertheless, it is included in the interests of comprehensive coverage of the literature on singing and health, and for a sense of balance in highlighting that negative health issues may be associated with singing where it is pursued professionally.

The study is structured in two parts with an initial qualitative enquiry through interviews followed by a larger-scale questionnaire survey. The nature and aims of the study are explained as follows:

'In the qualitative part of the present study, the following specific questions were addressed: (1) What problems do opera singers experience in their striving for artistic and technical perfection? (2) What type of coping strategies are used to overcome the problems? (3) What are the motivation factors that influence opera singers' professional activities? (4) What does a strong emotional singing experience mean for an opera singer? The aim was to gather data that would serve as material for the construction of a questionnaire. For the quantitative study, the focus was on singers' somatic problems, worry and about others' opinions, depressive tendencies, relation to their own voice, and health-promoting activities.' (2002:11)

Fifteen opera singers (8 female, 7 males) aged 27-65 were interviewed in the first stage. Common themes emerging were:

- Worry about negative criticism and ambivalence towards positive feedback
- A high level of self-criticism in relation to performance

- Worry about physical conditions (e.g. menstruation in the case of women) and problems (e.g. upper respiratory infections) that could impair the voice and a compulsive attitude towards testing the voice
- A sense that the body and mind work together in performance, but also that the body is separate and operates independently of the self
- A need for work-life balance, particularly in relation to family relationships and commitments
- In all of these areas, singers employed a range of coping strategies to try and avoid psychological, psychosocial and somatic problems

Sangren gives some fascinating quotations from singers in response to a request to describe 'memorable highly emotional singing experiences in front of an audience.' Gender differences were apparent, with men tending to emphasise the issue of 'mastery of technique' whereas women placed greater stress on a feeling of 'communication with the audience.' Notwithstanding these differences, the following accounts demonstrate vividly the intensity of the emotional experience, the sense of personal control, and the feelings of self-expression and satisfaction, which accompanied the use of the body and the vocal apparatus as a musical instrument:

*'What I find so fascinating, may be not so interesting for the audience, but for me subjectively, is when the instrument is functioning perfectly. It is great and terribly sensual, this feeling when you have the technique in your hand and you can do whatever you want. That the technique is obeying.'* (Male singer)

*'It is the need to express your self in music. It is so physical, you feel alive with your whole body, you are breathing. Expressing feelings like that in front of an audience, that is why I am singing.'* (Female singer) (2002:16)

In the second quantitative stage of the study, 49 professional opera singers (25 women, 24 men) completed a specially designed questionnaire to investigate 'psychological and somatic problems, health-promoting activities, and the relation to the voice.'

Out of 22 statements about worries, the voice and depressive tendencies rated on a five point scale (0-4), with several means above the mid-point of 2 for statements related to 'worry about others' opinions' and the singer's relation to their voice. The highest means being reported for: 'To be ill is very trying for my mental well-being' (2.40), 'When the voice does not function, the self esteem plummets. (2.39), and 'It is difficult to ignore negative criticism' (2.37). All the means for items related to 'depressive tendencies' are lower than 2, suggesting that in general depressive tendencies were not particularly prevalent. Unfortunately, however, Sandgren does not report standard deviations, or frequency distributions, so it is difficult to assess whether any singers had particular problems with depressive tendencies.

With respect to physical health problems (e.g. headache, respiratory infection), Sandgren is not explicit on how these were assessed (i.e. the nature of the rating scales or period of time singers were asked to consider). In addition, she does not report descriptive data for the sample, so it is not possible to determine how widespread or frequent somatic complaints were. Nevertheless, she reports that 'depressive tendencies' and 'somatic complaints' were significantly correlated.

Singers were asked to report on their 'health-promoting activities' (e.g. jogging, use of vitamins, relaxation exercises), but again no explicit details are given on the scales employed and no descriptive data are reported. However, a simple factor analysis was undertaken, and five factors identified. The first and most substantial factor 'represented techniques involving mental and relaxation exercises' with strong loadings from 'meditation', 'yoga', 'mental training' and 'relaxation exercises.'

As noted earlier, the focus of Sandgren's study is not directly relevant to the central concern of this review, but it has considerable value in highlighting the potentially negative aspects of singing with respect to wellbeing and health, in the context of the demands associated with being a professional opera singer.

### **4.3 Quantitative research studies using previously validated psychological measures**

Seven studies of singing were identified in which previously validated psychological measures were employed (see Table 13). Most attention will be given to four studies which consider the impact of singing on mood. The studies by Hills and Argyle (1998a, b) and Wise et al. (1992) are of limited interest and will be described only briefly.

#### **4.3.1 Singing and psychological mood states**

Four studies have measured mood states of participants before and after singing, using previously developed, standardised mood questionnaires (Rider et al., 1991; Valentine and Evans, 2001; Unwin et al., 2002; Kreutz, et al., 2004). In each case, the broad hypothesis under test is that singing will increase positive mood states and reduce negative mood states.

Among these studies, only Unwin et al. directly considers the qualities of the music being sung and the lyrics of the songs as potential factors influencing mood. The lack of such consideration in the remaining three studies is a significant limitation since on common sense grounds alone it might be expected that shifts in mood while singing would depend on the character and meaning of the music being sung.



Table 13: Studies using previously standardised psychological measures

Author /date	Sample	Singing Activity	Design	Profile of Mood States	UWIST Model Adjective checklist	Positive and Negative Affect Schedule	Oxford Happiness Inventory	Life Satisfaction Index A	Personal Orientation Inventory	Middleton Alienation Scale
Rider et al. 1991	17 musically trained people	A single ten minute session of solo singing of verses from ten familiar songs	All participants involved in four conditions - singing, toning, listening and control, with post test assessments	✓						
Wise et al. 1992	49 members of the community chorus, and 49 non-singing residents of the same community	Singing in a retirement village community chorus	Self completion questionnaire survey with comparisons between subgroups					✓	✓	✓
Hills and Argyle, 1998a,	230 adults from a university research panel. 71 belonged to music groups, particularly amateur choirs. Subgroups belonged to music groups, churches, both or neither	Singing in choirs	Self completion questionnaire survey with comparisons between subgroups				✓			
Hills and Argyle	275 adults recruited through local organisations and leisure groups. 76 belonged to music groups, particularly amateur choirs. Other leisure activities included church attendance, sport and watching TV soaps	Singing in choirs	Self completion questionnaire survey with comparisons between subgroups				✓			
Valentine, 2001	10 students having singing lessons, 13 members of a college choir, 10 members of the university swimming club	A single 30 minute singing lesson, a single 30 minute choral rehearsal	Three separate groups assessed before and after solo singing, choral singing and swimming		✓					
Unwin et al., 2002	107 members of the general recruited via advertisements. Half participated in a singing group (Exact N not given)	A short programme of two vocal exercises and five songs with non-English lyrics	Experiment with participants randomly allocated to either a singing group or a listening to music group	✓						
Kreutz et al., 2004	32 members of an amateur choral society	A single 60 minute rehearsal of Mozart's <i>Requiem</i>	Choir members assessed before and after a single rehearsal and a session listening to music			✓				

In the Unwin et al. study the potential impact of lyrics was 'controlled' by using non-English songs:

The choice of non-English lyrics was, in part, influenced by research indicating that the influence of lyric content may override musical accompaniment in altering the mood of listeners (Stratton and Zalonowski, 1994). Consequently, the non-English lyrics in this study aimed at minimising any association with the intended emotional content of the words. (2002: 178)

The songs to be sung were also carefully chosen to be in a major key and relatively simple and easy to learn:

The five songs, from Paraguay, Argentina, Africa and the Caribbean, were simple in structure, melody, tonality and rhythm. They were all in a major key and were either in simple duple (...) triple (...) or quadruple time, within the range of an octave (...) with a simple structure, and between 8 and 16 bars in length. (2002: 178)

In the Rider et al. study, participants 'sang one or two verses of ten familiar songs while they listened to piano accompaniment.' The songs were in comfortable keys for singing, but little significance was attached to the words since 'if subjects could not remember verses, they were told to repeat familiar verses.' (1991: 77). Valentine and Evans give no information at all on the repertoire sung by the solo and choral singers. In the Kreutz et al. study, the piece being rehearsed was Mozart's *Requiem* – a highly charged emotional work, in a minor key - which might be expected, for some people at least, to generate a mood of reflective sadness rather than a more positive mood. The *Requiem* is sung in Latin, and the meaning of the words may or may not have had significance for the singers.

Dario et al. (2006) is the only experimental study under review which manipulated the emotional character of sung material in order to assess the impact of 'happy' and 'sad' songs, but their investigation did not employ a standardised mood assessment instrument to determine whether singing happy or sad songs generated a positive or negative mood.

The four studies are difficult to compare as three different mood instruments are employed: two studies use the Profile of Mood States (POMS) (Rider et al., Unwin et al.), one the UWIST Mood Adjective Checklist (Valentine and Evans), and one the Positive and Negative Affect Scale (PANAS) (Kreutz et al.)

### **Rider et al. (1991) and Unwin et al. (2002) - Profile of Mood States (POMS)**

Unwin et al. provide a fairly detailed description of POMS:

The POMS is a rapid, inexpensive self-report method of assessing mood or affective states. It measures six identifiable mood or affective states: Tension-Anxiety; Depression-Dejection; Anger-Hostility; Vigour-Activity; Fatigue-Inertia and Confusion-Bewilderment. Most participants can complete the POMS in 3-5 minutes. Although originally designed for a clinical population, it is also used extensively in healthy populations, has been used successfully to support the observations of physiological changes associated with mood change (Gibson, 1997; Chlan, 1995) and as a measure of mood change following various forms of exercise... (2002: 177)

The POMS comes in a number of different versions and it is not clear which version was used in either paper. The longer version consists of 65 items each scored on a five point scale from 'Not at all' to 'Extremely', while the short version consists of 30 items (5 per scale). As the mean values reported in both papers for the negative scales are generally very low, and as Unwin et al. state that the scale can be completed within 3-5 minutes, it is assumed that both sets of researchers employed the short form.

Rider et al. compared the impact on mood of singing, toning, listening to a recording of Gregorian chant and silence with pre and post assessments. As the pre-test scores for the four conditions did not vary, they report means and standard deviations for post-test scores only, and make statistical comparisons across the four conditions. The data show that for all groups at post-test, measured tension, depression, anger, fatigue and confusion was very low while vigour appeared to be high. For all of the scales no significant main effects emerged across the four conditions. As no comparisons were made between pre and post test scores, it is not known whether any shifts occurred in response to some or all of the conditions. It has to be concluded, therefore, that the Rider et al. study tells us little about the impact of singing on mood.

Unwin et al. compared the impact on mood of involvement in group singing and listening to group singing with pre and post assessments using POMS. They report data graphically rather than in tabular form for only 5 of the 6 scales (the Depression scale was excluded due to varying results when parametric and non-parametric tests were employed), so mean values can only be estimated and no information is given on standard deviations. Judging from the bar charts presented (Figures 1 and 2, p. 181), at pre-test both the singing and the listening groups had very low means on the negative mood states – tension, anger, confusion and fatigue – and relatively higher means on the positive mood measure of vigour.

Unwin et al. report that statistically significant shifts occurred for both singing and listening groups on all measures, with negative moods reduced and vigour increased, but no significant difference emerged between the two groups 'indicating that the mood of both singers and listener groups changed in a similar manner over time' (2002: 180). Somewhat inappropriately, however, given the lack of statistical support, Unwin et al. go on to highlight a trend in favour of group singing:

Although group effects were not significantly different on any of the statistical tests, examination of group means and graphical displays indicate a consistent trend favouring bigger effects for the singing group (Figures 1 and 2). Note that the values for tension and fatigue decrease more for singers than listeners and values for vigour increase more for singers than listeners. (2002: 180)

Despite the considerable difference in design and procedure between these two studies, the fact that one is concerned with individual singing the other with group singing, and the problems of data presentation, they are in fact

consistent in showing no difference in the effects of singing and listening on mood states as measured by the POMS.

However, neither study pays any attention to the psychological meaning of the average profiles obtained, nor, more importantly, to the meaning of changes occurring from pre to post test, and this is undoubtedly a serious weakness in both studies. To illustrate this point, consider the results reported by Unwin et al. on the 'Anger-Hostility' scale for the listener group. The listeners appear to have a mean score of about 6 at pre-test for 'Anger-Hostility' and a mean of 3 at post-test, and both of these values must be considerably below the mid-point on the scale. This point to very low levels of 'Anger-Hostility' at the start of the experiment and even lower levels just 30 minutes later – indeed, given the rating scale employed, many of the participants must have said that they had no angry feelings at all at the start of the experiment, and no such feelings at its conclusion. There is also a distinct likelihood that the Unwin et al. study was highly vulnerable to the operation of substantial demand characteristics. Both the purpose of the measuring instrument and the experiment itself must have been very obvious to the participants and it is not difficult to imagine that they would have obliged the experimenters by reporting less negative and more positive moods following singing than they had before hand!

It has to be seriously questioned, therefore, whether either study has any scientific credibility or relevance to the issue of whether active participation in singing has benefits for wellbeing or health.

### **Valentine and Evans (2001) – UWIST mood adjective checklist**

Valentine and Evans (2001) report a comparison of the effects of solo singing, choral singing and swimming on mood as measured by the UWIST mood adjective checklist. They offer the following account of how this instrument works:

The UWIST mood adjective checklist was chosen to measure the three main dimensions of mood in a normal sample. This has been shown to have satisfactory psychometric properties (...). Three scales measuring tense arousal, energetic arousal and hedonic tone respectively, each contain eight items, four positively and four negatively weighted, rated on a 4-point scale. (2001: 116)

Samples of solo singers, choral singers and swimmers, completed the checklist before engaging in singing or swimming for 30 minutes, and then completed the checklist again. The results obtained are reported in Table 14.

Table 14: Results from two studies on singing which measure mood changes

Author / date	Sample	Condition	Measure employed Mean (SD)		Statistical analysis	Effect size post to pre Cohen's d								
			UWIST Mood Adjective Checklist Before	Checklist After										
Valentine 2001	Students having singing lessons (N=10), Members of a college choir (N=13), Members of a swimming club (N=10)	Singing alone	Tense arousal		Significant reduction in tense arousal across all groups (F=13.93, p<0.001), but no significant interaction effects were found.	-0.63								
			17.00 (6.60)	13.60 (3.75)										
		Choral singing	Energetic arousal				Significant increase in energetic arousal across all groups (F=58.65, p<0.001), and a significant interaction between time and activity (F=10.21, p<0.001), with increases greater for swimming than either singing activity	-0.41						
			16.58 (5.86)	14.53 (3.93)										
		Swimming	Hedonic tone						Significant increase in hedonic tone across all groups (F=57.44, p<0.001), and a significant interaction between time and activity (F=5.79, p=0.007), with increases greater for swimming than either singing activity.	-0.60				
			19.60 (5.44)	12.80 (3.26)										
		Singing alone	Tense arousal								Significant interaction between time and condition (F = 6.42, p < 0.02). Follow-up tests: positive affect increased significantly after singing (p<0.05), but not after listening.	0.75		
			23.20 (4.52)	26.30 (3.71)										
		Choral singing	Energetic arousal										Significant main effects for condition and time and an interaction (F = 145.91, p < 0.001). Post hoc tests: significant decrease in negative affect after singing (p<0.05) and a significant increase after listening (p<0.05)	0.79
			20.92 (4.99)	24.77 (4.85)										
		Swimming	Hedonic tone											
16.40 (4.03)	27.60 (3.34)													
Singing alone	Tense arousal		Significant interaction between time and condition (F = 6.42, p < 0.02). Follow-up tests: positive affect increased significantly after singing (p<0.05), but not after listening.	1.05										
	25.00 (4.39)	29.10 (3.35)												
Choral singing	Energetic arousal				Significant main effects for condition and time and an interaction (F = 145.91, p < 0.001). Post hoc tests: significant decrease in negative affect after singing (p<0.05) and a significant increase after listening (p<0.05)	0.57								
	25.00 (4.22)	27.38 (4.07)												
Swimming	Hedonic tone						Significant main effects for condition and time and an interaction (F = 145.91, p < 0.001). Post hoc tests: significant decrease in negative affect after singing (p<0.05) and a significant increase after listening (p<0.05)	1.94						
	23.40 (4.27)	30.10 (2.38)												
Kreutz 2004	Members of an amateur choral society (N= 32)	Choral singing							Measure employed Mean (SD)		Significant interaction between time and condition (F = 6.42, p < 0.02). Follow-up tests: positive affect increased significantly after singing (p<0.05), but not after listening.	0.50		
									Before	After				
		Listening to music							PANAS				Significant main effects for condition and time and an interaction (F = 145.91, p < 0.001). Post hoc tests: significant decrease in negative affect after singing (p<0.05) and a significant increase after listening (p<0.05)	-0.08
									2.86 (0.51)	3.15 (0.64)				
		Choral singing							Positive affect					
			2.85 (0.67)	2.79 (0.81)										
		Listening to music	Negative affect						Significant main effects for condition and time and an interaction (F = 145.91, p < 0.001). Post hoc tests: significant decrease in negative affect after singing (p<0.05) and a significant increase after listening (p<0.05)	3.44				
			1.31 (0.40)	1.18 (0.24)										
		Listening to music	PANAS		Significant main effects for condition and time and an interaction (F = 145.91, p < 0.001). Post hoc tests: significant decrease in negative affect after singing (p<0.05) and a significant increase after listening (p<0.05)	3.44								
			1.23 (0.25)	2.20 (0.31)										

Rider et al. (1991) Only post-test data are reported on the six POMS scales. No significant main effects are reported for tension, depression, anger or vigour. A significant effect emerged for fatigue and post hoc analysis revealed that the listening condition was significantly lower in fatigue than the control condition. No evidence, therefore, that singing had measurable and significant impacts on any dimension of mood compared with other conditions.

Unwin et al. (2002) Pre, post and follow-up data on means for 5 of the 6 POMS scales are reported in the form of bar charts, rather than directly in tables. In addition, no values for standard deviations are given and consequently effect sizes cannot be computed.

It is clear that solo singing, choral singing and swimming all had a statistically significant impact on mood – reducing tense arousal, and increasing energetic arousal and hedonic tone. The figures reported show clearly that the changes are very similar for both solo and choral singing, but tend to be more marked for swimming (and significantly so in the case of energetic arousal and hedonic tone). Table 14 gives a measure of effect size for the activities on mood. It is clear that Cohen *d* values are similar for tense arousal – and almost identical for solo singing and swimming, with the effect of choral singing being lower. For energetic arousal solo and choral singing appear to have a fairly strong effect, but clearly nowhere near as strong as the effect of swimming. The effect sizes for hedonic tone show a moderate effect of choral singing, and larger effects for solo singing and swimming.

As with the Rider et al. and Unwin et al. studies, it is important to look carefully at the actual values reported for mood assessments. Given the construction and scoring of the scales, a range of values from 8-32 is generated, with a mid-point on the scales of 20. All the means reported are close to the mid-point, and the changes from pre to post-test represent average shifts of approximately one point on the four-point rating scale employed for individual items in each scale. For tense arousal, for example, all group means are below the mid-point at pre-test and fall still lower at post-test. On average, participants in the experiment will at pre-test have given ratings of 2 on the individual items making up this measure, and then tended towards giving ratings of 1-2 at post test.

The scale employed appears to be more appropriate than the POMS scale used by Rider et al. and Unwin et al., but the findings may still have been influenced to an unknown degree by demand characteristics, as both the purpose of the experiment and the function of the mood checklist would have been clearly apparent to participants. The possible role of demand characteristics is recognised by Valentine and Evans, but they suggest, without further elaboration, that this factor ‘would not explain interactions between time and activity.’ They also acknowledge that participants were not ‘randomly assigned to treatment conditions, but argue that the skilled nature of the activities considered meant that it was not feasible or realistic to randomly allocate participants.

In conclusion, this study appears to show that choral and solo singing, have similar positive effects on mood, but that the impacts of swimming, as a more energetic physical activity, appear to be stronger. Whether such shifts in mood have a bearing on broader dimensions of wellbeing and health, however, is an open question. Valentine and Evans acknowledge that their study is preliminary and call for further research:

These data should be regarded as preliminary, particularly in view of the small sample sizes. However, given the pandemic nature of anxiety and depression (...), the fact that most advances in public health come from prevention rather than treatment (...), and the evidence for the effect of positive emotional states on physical wellbeing (...), it is imperative that further analytic investigations are

undertaken into the mechanisms underlying the positive effects of engaging in recreational activities, particularly given the current increases in leisure time and in aging populations. (2001:119)

### **Kreutz et al. (2004) – Positive and Negative Affect Scale (PANAS)**

Kreutz et al. (2004) report a study of an existing amateur choir in the process of rehearsing Mozart's *Requiem*. They asked members of the choir to complete the Positive and Negative Affect Scale (PANAS) before and after a standard rehearsal, and invited them to return one week later to simply listen to a recording of the *Requiem* completing the same scale before and after listening. Kreutz et al.'s description of singing was quoted in an earlier section, and commended as a relatively full characterisation of the activity under study. It is appropriate to give their description of the listening condition at this point:

During the second session 1 week later, the pieces from Mozart's *Requiem* were presented from CD, and articles on singing from an eighteenth century encyclopedia of the arts (...) were read aloud. Participants were seated during the entire session. When the music played, they were instructed to listen to the music attentively *as if* they were engaged in singing. Moreover, it was ensured that listening to speech and music under the listening condition had the same proportion as singing and listening to the conductor under the singing condition.

Again, the detail is admirable, although purists might wonder why, given the care to ensure an equal proportion of listening to speech and music, the researchers did not also instruct participants to stand and sit while listening in the same proportions as while singing.

Interestingly too, as a point of omission, the researchers fail to tell us whether the members of the choir also had their scores in front of them while listening. Presumably they did, as they would undoubtedly have had scores during the previous week's rehearsal. An important point to highlight here is that participants were asked 'to listen to the music attentively *as if* they were engaged in singing'. It is quite possible that most of the choristers would have followed the scores and mouthed the words, and some may even have sung *sotto voce*. It should be remembered that when a person engages in group singing, they are both producing musical sounds and listening to the sounds produced by others around them, and these sounds in themselves may contribute to any impact on mood or wellbeing which group singing may have. This discussion makes explicit that the comparison involved in this study is concerned with whether any additional impact on mood state comes from *the vocal production of musical sounds per se*.

Despite the admirable detail with which Kreutz et al.'s describe the two 'treatments' in their design, it can be argued that even more detail is needed to be clear about the nature of the activity and experience of the choral singers, when singing and listening!

Kreutz et al. are also detailed in their account of the PANAS instrument:

The PANAS consists of 20 items, 10 representing positive affect (e.g. “I feel fine”), and 10 items representing negative affect (e.g. “I feel depressed”). Participants were asked to mark each of the items on a scale from 1 (“very little or not at all”) to 5 (“extremely”) according to their current feeling. The PANAS was filled in once again at the end of each session. (2004: 626)

It might be assumed that people who give a high rating in response to “I feel fine” would tend to give a low score in response to “I feel depressed” and vice versa. Empirically, however, positive and negative affect emerge as separate variables, rather than being two ends of a single spectrum. Kreutz et al. therefore report mean scores for positive and negative affect. It can be seen from the results reproduced in Table 14 above that the scores reported are not summed scores, but means of the 10 item ratings making up each scale. This is helpful, as the scores can be interpreted directly in terms of the anchors given (very little or not at all – extremely).

The results obtained for positive affect show that at pre test, means were very close to the mid-point with standard deviations around 0.5-6, indicating that most people gave scores of 2, 3 or 4 on the items initially. Following the intervention, the mean increases for the singing group, but remains effectively unchanged within the listening group, and while the effect size on this scale for the singing group is moderate, the mean and standard deviation indicates that participants have shifted their ratings by no more than one point on the scale.

The results obtained for negative affect are more interesting. Here, the means at pre test are very low with low standard deviations indicating consistently low levels of negative affect in the group prior to singing and listening. Following the intervention, the mean for the singing group falls even lower, with again a moderate effect size, but the mean for the listening group increases, with a very strikingly high effect size. This suggests that listening resulted in a marked effect on negative mood state, although the mean rating is still below the mid-point on the scale.

In conclusion, the results show that singing has a significant, but small impact on mood – with mood slightly more positive and slightly less negative. More significantly, however, they reveal that asking people to attend a regular rehearsal slot to listen to a recording of singing, rather than sing themselves, results in no measurable change in positive mood, but rather dramatically increases their negative mood state!



### **4.3.2 Musical experiences and happiness**

#### **Hills and Argyle (1998a,b)**

As described in the previous section, Hills and Argyle report two studies in which members of 'musical groups, particularly amateur choirs' completed a specially constructed questionnaires to assess 'the intensity of their emotional feelings ... for musical activities' (1998a:91). In addition, participants were asked to complete the Oxford Happiness Inventory – a 29-item scale to measure 'happiness' which is claimed to have 'high reliability and satisfactory validity. In the 1998a paper, the hypothesis that 'musical experiences' are 'associated with happiness' was tested by correlating the 'mean item scores' on the musical experience scale with the OHI score, and a significant positive correlation was found supporting the hypothesis ( $r=0.15$ ,  $p<0.05$ ), but clearly it is very weak.

Hills and Argyle's interpretation of this finding is that the OHI is a 'broad and eclectic measure of happiness covering many different domains' whereas musical experiences are 'quite specialised.' They argue on this basis that musical experiences could make a significant contribution to 'some specific domains of the OHI which, nevertheless, would be too limited to have any significant effect on the overall happiness score.' Following this line of argument they factor the OHI and report seven factors. These are then considered in relation to the five factors found for the musical experiences scale (described in Table 8 above) and significant positive correlations emerge for three OHI and musical factors pairings. Hills and Argyle conclude that these findings 'support the hypothesis that in detail, if not in general, musical (...) experiences are positively associated with happiness' Hills and Argyle (1998a:99).

The same analytical strategy is adopted in the 1998b paper with similar results. The mean affect score on a reduced musical experience scale is positively but weakly correlated with the overall OHI score ( $r=0.23$ ,  $p<0.001$ ), and significant correlations are found with three of the seven OHI factors.

The Hills and Argyle studies do have significant weaknesses as noted earlier – not least the lack of specificity regarding participants' musical experiences and failure to specify the sex composition of the sample. In addition, the Oxford Happiness Questionnaire (a further development of the Inventory) has been the subject of a damning critique by Kashdan (2004), which seriously questions its validity as a measure of well-being.

### **4.3.3. Choral singing and successful aging**

#### **Wise et al. (1992)**

In one of the earliest studies identified for this review, Wise et al. asked 49 members of 'a retirement village community chorus' to complete three questionnaires to assess 'life satisfaction, alienation and self-actualization.' A further sample of 49 community members who were not members of the choir

was also recruited and completed the same questionnaires. The authors provide a demographic profile for the two groups, which appear to indicate that they were broadly comparable in terms of age, class, education, religion and self-assessed health, but strangely they neglect to report the sex composition of their samples. No differences were found in mean life satisfaction or alienation scores between the two groups, but the choral group, perhaps surprisingly, had lower scores on two of the three self-actualization sub-scales of 'inner-directedness' and 'actualization.' The authors conclude that 'group musical activity provides a link to previous satisfying experiences' and may help with a process of 'successful aging', and that 'the nature of the activity' may require 'a willingness to give up some inner directedness' – hence the lower scores found for the choral group on two of the scales. However, the authors acknowledge weaknesses in the design of their study and limitations in the measures used.

#### **4.4 Quantitative studies that use objective biological/physiological measures**

Eight studies in the corpus reviewed employed objective biological/physiological measures in assessing the impacts of participation in singing. Details of these studies are given in Table 15.

Most of these measures serve as physiological or endocrine markers of arousal or individual reactions to stress (e.g. cortisol, heart rate, blood pressure). Predictions of changes are made in accordance with whether the singing experience is seen to be physically arousing, emotionally stimulating, or stressful on the one hand or relaxing and calming on the other. Immunoglobulin A, an antibody found in saliva and other mucosal membranes is taken to be a marker of immune system functioning, which is also regarded as responsive to stress. Oxytocin is a hormone considered to underpin feelings of group bonding, attachment and comfort and is assumed to reflect such experiences in response to singing.

In reviewing these studies, detailed attention will be limited to the most commonly used measure, sIgA.

Table 15: Studies of the effects of singing using objective biometric measures															
Date	Author	Sample	Singing Activity	Design	Salivary Igmoglobulin A	Cortisol	Oxytochin	TNF-alpha	Prolactin	Heart rate	Heart rate variability	Blood pressure	Electromyographic temsion	Peripheral skin temperature	Skin conductance
1991	Rider et al.	17 musically trained people	A single ten minute session of solo singing of verses from ten familiar songs	All participants involved in four conditions - singing, toning, listening and control, with post test assessments	✓					✓			✓	✓	
2000	Beck et al.	31-34 members of a semi-professional chorale	Two 150 minute rehearsals (one early and one late) and one 90 minute performance of Beethoven's <i>Missa Solemnis</i>	Choir members assessed before and after 2 rehearsals and one performance	✓	✓									
2001	Valentine	10 students having singing lessons, 13 members of a college choir, 10 members of the university swimming club	A single 30 minute singing lesson, a single 30 minute choral rehearsal	Three separate groups assessed before and after solo singing, choral singing and swimming						✓		✓			
2002	Kuhn	33 undergraduate university students	A single 30 min session of call and response singing, and music making with instruments	Participants randomly assigned to three conditions, singing, listening and no activity control, with pre and post test assessments	✓										
2003	Grape	8 professional and 8 amateur solo singers	A single 60 minute singing lesson	Participants assessed before and after the lesson		✓	✓	✓	✓		✓				
2004	Kreutz et al.	32 members of an amateur choral society	A single 60 minute rehearsal of Mozart's <i>Requiem</i>	Choir members assessed before and after a single rehearsal and a session listening to music	✓	✓									
2006	Beck et al.	10 solo university student singers	A series of single lessons and some performances	Participants assessed before and after singing lessons and performances	✓	✓									
2006	Galati et al.	33 professional or semi-professional solo singers	Two 2-3 minute sessions singing three sad melodies and three happy melodies (all specially composed without lyrics)	Participants monitored before, during and after singing											✓
					5	4	1	1	1	2	1	1	1	1	1

Details of time and repertoire noted if given by the author(s)

#### 4.4.1 The impact of singing on salivary immunoglobulin A

Five studies have employed measures of salivary immunoglobulin A (sIgA) in assessing the physiological impact of singing. Table 16 gives the details of these studies, three of which are concerned with choral or group singing, and two with individual singing.

Technical details of how researchers assayed sIgA are given in footnotes to the table. These details are important in understanding what the figures reported actually mean. One aspect of these results which is immediately obvious, and potentially problematic, is the variation in nature and value of the results obtained across these studies. This is particularly striking in respect of the two studies by Beck and his colleagues.

In order to provide a common basis of comparison across the studies, Cohen *d* values were calculated as a measure of effect size. Cohen *d* is the difference between means expressed in relation to pooled standard deviation values. Common interpretations of Cohen's *d* are that 0.2 = small effect; 0.5 = medium effect, and 0.8+ = large effect. In the case of Beck et al. (2000), Kuhn (2001), Kreutz et al. (2004) and Beck et al. (2006), estimates of effect size were based on changes from pre to post test for each condition in the study. In the case of Rider et al. (1991) only post test scores were reported (as no significant differences in sIgA levels were apparent at pre test), and the estimates of effect size are based on comparisons between the experimental and control conditions.

Four of the studies report significant increases in sIgA in response to singing (as compared, for example, with listening to a recording of singing (Kreutz et al., 2004) or sitting in silence (Kuhn, 2001)). The exception is the study by Rider et al. (1991), in which singing was associated with lower sIgA levels than either silence or listening (since pre test values are not reported it is not possible to say whether the values for singing represent a fall from the baseline or no increase).

Interesting patterns appear to emerge from considering effect sizes. Firstly, active participation in singing produces a larger effect on sIgA than listening to singing (Kuhn, 2001; Kreutz et al., 2004). Secondly, performance produces a stronger effect on sIgA than rehearsal (Beck, et al., 2000, 2006). Thirdly, there is some suggestion that a short session of singing popular songs in a small group can have a similar effect on sIgA, as a lengthy performance of a major choral masterpiece – Beethoven's *Missa Solemnis*!

Such comparisons and interpretations need to be treated with considerable caution of course. The value of careful attention to the details of existing studies is that it offers guidelines for future studies, which might explore more systematically the potential impacts of singing on immune function.

Table 16: Results from five studies on singing which measure salivary immunoglobulin A

Author / date	Sample	Condition	Measure employed Mean (SD)		Statistical analysis	Effect size post to pre Cohen's d <sup>b</sup>
			Before	After		
Beck et al. 2000	Semi-professional chorale (N = 31-34)	Rehearsal one	Ratio s-IgA/protein <sup>1</sup> 0.040 (0.042)	0.060 (0.048)	Pre-post: t = 2.98, p < 0.004	0.44
		Rehearsal two	0.050 (0.044)	0.100 (0.080)	Pre-post: t = 4.81, p < 0.0001	0.77
		Performance	0.036 (0.030)	0.073 (0.049)	Pre-post: t = 5.87, p < 0.0001	1.06
Kuhn 2001	Undergraduate students (N = 26)	Active participation (N = 10)	4.24 (1.02)	5.93 (1.83)	ANOVA main effect for groups, F not given, p < 0.01. Post hoc analysis: active group shows greater increase than passive and control groups (p < 0.01)	1.14
		Passive listening (N = 9)	3.89 (1.30)	4.28 (1.55)		
		Control (N = 7)	4.67 (1.32)	4.24 (1.80)		
			s-IgA precipitin ring diameter (mm) <sup>2</sup>			
Kreutz et al. 2003	Amateur choral singers (N = 31)	Rehearsal	Ratio s-IgA/albumin <sup>3</sup> 3.66 (3.15)	5.28 (5.26)	ANOVA main effect for condition, F = 10.41, p < 0.005; significant interaction between condition and time, F = 4.32, p < 0.05. Post hoc analysis: s-IgA levels higher after singing, p < 0.005, but not listening	0.37
		Listening to singing	4.10 (4.20)	4.49 (3.78)		
Beck et al. 2006	Student solo singers (N = 10)	Rehearsal	Ratio s-IgA/protein <sup>4</sup> 1148.28 (582.15)	1188.83 (547.98)	Pre-post: t = 1.596, n.s.	0.07
		Performance	1162.80 (653.16)	1365.30 (784.49)	Pre-Post: t = 3.125, p = 0.012	0.28
Rider et al. 1991	Musically trained participants (n = 17)	Control (silence)	s-IgA concentration mg/ml <sup>5</sup> 43.35 (30.69)		ANOVA main effect, F = 2.87, p < 0.05. Post hoc analysis: singing lower s-IgA levels than silence and listening, but not toning	--
		Listening to music	47.23 (26.79)			
		Toning	36.16 (18.83)			
		Singing	28.54 (29.93)			

Table 16 Footnotes – details of the assaying of s-IgA

1. 'A problem in using saliva for measuring concentrations of protein is how to overcome the well-known inverse relationship between the amount of saliva produced in a given period (flow rate) and the concentration of the s-IgA protein. Our method controlled for flow rate by also measuring the total amount of protein in the saliva. Thus changes in concentrations of s-IgA or cortisol are expressed as proportions of total protein. Laboratory analyses of the proteins were conducted according to standard procedures recommended by manufacturers of the tests. Test plates contain a layer of gel embedded with anti-IgA. Saliva samples are loaded into the wells and incubated at room temperature for 48 hours. The IgA diffuses from the well into the gel. The resulting diameter is read with a microscope and converted by a calibration curve to weight (mg/dL). The normal range is 3-25 mg/dL. (pp. 93-4)
2. 's-IgA concentrations were determined by Radial Diffusion assay. This procedure is based on the phenomenon of an antigen-antibody reaction in a gel medium forming a precipitin ring. Measured samples of equal volume (5 ml) of the test saliva were delivered into a well cut into the commercially available, pre-prepared plates. The area within the precipitin ring is proportional to the antibody concentration. Effects on the immune system were determined by calculating the difference in the precipitin ring diameters of the before and after samples of the same subject.' (pp. 34-5)
3. 'Measured parameters in saliva samples were immunoglobulin A, albumin, and cortisol. Albumin levels served both as an exclusion criterion for blood contaminated saliva samples and for correcting the s-IgA measures for effects of saliva flow density. (...) saliva was analysed for concentrations of s-IgA and albumin by use of a fully automated nephometric analyses. The assay protocol has been adapted to the expected range for saliva concentrations of s-IgA between 0 and 120 mg/dL and albumin (0 – 27 mg/dL). S-IgA/Albumin is without unit because the units for both parameters are identical (mg/dL).' (pp. 627-9)
4. The saliva samples were assayed using an ELISA procedure. Sandwich ELISA for s-IgA was performed with antibodies from Research Diagnostics. S-IgA concentrations were calculated by interpolation from a standard curve. To standardize the results of the ELISA, a quantitative analysis of the protein present in each person's saliva was necessary (s-IgA (µg/ml)/Whole protein (µg/ml)).' (pp. 6-9)
5. 'The measurement of s-IgA was intended to provide a measurement of immune functioning with greater concentrations of s-IgA correlated with higher immunocompetence and greater relaxation. From timed saliva collections, 5 ml samples were removed from which s-IgA was assayed using standard radial immunodiffusion techniques. Raw data consisted of posttest scores in mg/dL.' (pp. 76-7)
6. Cohen d gives an estimate of 'effect size' in which the difference between means is expressed in relation to pooled standard deviation values. Common interpretations of Cohen's d: 0.2 = small effect; 0.5 = medium effect; 0.8+ = large effect.

## Limitations of the studies of sIgA

Each of the studies assessing sIgA has its limitations, not least with respect to the character of the singing investigated and the sizes of the samples involved.

A more fundamental problem, however, is the difficulty of determining whether the kinds of changes in sIgA reported have any significance whatsoever for the wellbeing and health of the people involved. On the face of it, an increase in an antibody said to be 'part of the body's frontline defences against infection' would appear to be of potential value to health. There are, however, significant limitations to this kind of evidence in relation to health.

- Firstly, these studies are reporting shifts in average values over periods ranging from a few minutes to a couple hours, and we have no data on how long such changes were sustained after singing.
- Secondly, no evidence is presented to show that the level of changes have any real clinical significance with respect to an individual's resistance to infection.

- And thirdly, and most tellingly, the change may have come about for reasons entirely unconnected with the person's activity of singing –being in close proximity with a group of people who are breathing in and out energetically. In other words, the heightened immune response may be a direct response to an increased risk of respiratory infection.

#### **4.4.2 The impact of singing on additional biomarkers**

Table 17 summarises findings for additional biomarkers employed in the eight studies considered in this section.

As noted above, most of these measures, with the exception of oxytocin, can be regarded as measures of stress-relaxation. The key example is the hormone cortisol, which has been assessed in four studies with mixed results. The remaining measures are employed in only one or two studies, again with mixed results. Overall, the findings appear to show simply that singing is an activity associated with a range of physiological changes indicative of arousal or mild stress – a conclusion which is hardly surprising. Interestingly, no studies in this set make a clear distinction between positive experiences of challenge which are motivating i.e. 'eustress' and negative experiences of 'distress' which arise when external demands exceed available resources for coping.

The finding that oxytocin increased in both amateur and professional singers in the course of a singing lesson is an interesting finding, and points towards a different kind of physiological process at work during singing. Oxytocin is a hormone associated particularly in women with childbirth and breast-feeding and in both sexes with processes of interpersonal intimacy and bonding. The increase in oxytocin points to enhanced feelings of wellbeing associated with feelings of connectness with others, and may taken place even more strongly in the context of group singing – a hypothesis which remains to be tested in future research.

Table 17: Summary of findings from studies using biomarkers other than sIgA

Cortisol	A hormone associated with arousal, stress and energy mobilisation (Grape, et al., 2003). 'Increased cortisol levels have been related to a wide range of stressful life experiences...' (Beck, et al., 2000)	<u>Decreased</u> after two and half hour choral rehearsals. <u>Increased</u> after a one and a half hour choral performance (Beck, et al., 2000) <u>Increased</u> for men after a 60 minute individual singing lesson. <u>Decreased</u> for women (Grape, et al., 2003). <u>No change</u> after a 60 minute choral rehearsal (Kreutz, et al., 2004). <u>No change</u> after individual singing rehearsals or performances, (Beck et al., 2006).
Oxytocin	A pituitary hormone released in women during childbirth and breastfeeding. Important in both sexes in relation to attachment and bonding (Grape, et al., 2003).	<u>Increased</u> in individual amateur and professional singers after a 60 minute singing lesson (Grape, et al., 2003)
TNF-alpha	Tumor Necrosis Factor alpha is a cytokine which increases in inflammatory conditions and also in response to stress (Grape, et al., 2003).	<u>Increased</u> in professionals after a 60 minute individual singing lesson. <u>Decreased</u> in amateurs (Grape, et al., 2003)
Prolactin	A hormone which increases during stressful situations involving a lack of control over circumstances (Grape et al., 2003)	<u>Increased</u> for men after a 60 minute individual singing lesson. <u>Decreased</u> for women (Grape, et al., 2003).
Heart rate	A simple measure of short-term autonomic arousal or stress – higher arousal leading to higher heart rate (Rider, et al., 1991).	<u>No change</u> after 30 minutes of group or solo singing (Valentine and Evans, 2001) <u>Increased</u> after 10 minutes of singing compared with a no activity control condition (Rider, et al., 1991)
Heart rate variability	A marker of sympathetic nervous system activation. Measured in terms of total, high and low frequency power. Indicative of 'cardio-physiological fitness' (Grape, et al., 2003).	<u>Increased</u> power for professional singers after a 60 minute singing lesson. <u>No change</u> for amateurs (Grape, et al., 2003)
Blood pressure	A measure of short-term arousal and stress. Stress leading to raised blood pressure.	<u>No change</u> in systolic or diastolic blood pressure after 30 minutes of group or solo singing (Valentine and Evans, 2001)
Electromyographic tension	A measure of muscular tension. Greater tension indicative of activity and stress.	<u>No difference</u> after 10 minutes of singing compared to no treatment control condition (Rider, et al., 1991)
Peripheral finger temperature	A simple measure of blood supply to the skin. Less blood supply, indicative of stress, results in lower skin temperature.	<u>No difference</u> after 10 minutes of singing compared to no treatment control condition (Rider, et al., 1991)
Skin conductance	A measure of the moistness or dryness of the skin. Moist skin, indicative of stress, results in increased electrical conductance.	<u>Increased</u> in individual singers after four minutes of singing specially composed 'happy' and 'sad' melodies (Galati, et al., 2006)



## **4.5 Quantitative studies that use standardised physical and mental health measures**

Only two studies in the entire corpus have undertaken experimental and controlled investigations of the effects of singing using standardised health measures (Houston et al., 1998; Cohen et al., 2006, 2007). In both cases the samples involved were of elderly people who were assessed before and after a structured programme of singing. Appropriate control groups were also assessed. Both studies offer evidence that involvement in singing had statistically significant effects on health. Findings are also considered to be substantively significant, particularly given the advanced age of the people involved.

Given the nature of the designs and measures employed in these studies and the positive outcomes reported, it is appropriate to give close attention to the results reported. The following critique is informed by recent discussions on the quality of health research and the use of statistics by Sterne and Davey Smith (2001) and Ioannidis (2005).

### **4.5.1 Effects of singing groups in residential care homes**

#### **Houston et al. (1998)**

Houston et al. (1998) examine the effects of an 'old time sing-a-long' designed as 'a laughter-inducing social/diversional activity' on the psychological wellbeing of older people in residential care. The specific hypothesis was that 'a group of residents receiving a structured intervention of an old time sing-a-long would have lower anxiety and depression than a group of residents receiving standard residential care.' (1998:329)

Sixty-one residents in 6 homes participated in the study, 31 in the intervention and 30 in the control conditions. Their average age was 83.7 years. The hour-long singing sessions took place once a week over four weeks. Participants were assessed using the General Hospital Questionnaire-28 (GHQ-28), measuring 'somatic symptoms', 'anxiety and insomnia', 'social dysfunction' and 'severe depression', and the Hospital Anxiety and Depression Scale (HADS), measuring 'anxiety' and 'depression'.

Houston et al. adopt stringent standards in the statistical analysis of their data:

- Firstly, they show that the intervention and control groups did not differ at baseline (using t-tests) and that no significant differences were apparent across the homes involved in the study (by ANOVA)
- Secondly, they use analysis of covariance to take account of pre-test scores as covariates in assessing whether the intervention had any impact.
- Thirdly, they take into account that multiple hypotheses were being applied (6 in all) and lower the alpha value for rejection of the null hypothesis to  $p < 0.008$ .
- Fourthly, attention is given in a footnote to issues of power and the effects sizes associated with the intervention.

The findings reported by Houston et al. are reproduced in Table 18.

The apparent changes induced by the intervention are impressive. It appears, for example, to have halved the GHQ and HADS anxiety scores, and reduced by approximately 40% the score on the HADS depression scale. The intervention had no impact on the GHQ 'severe depression' score, however, but this is not commented on, and no effects were found for somatic symptoms and social dysfunction. Overall, therefore, the pattern of results points to beneficial effects from the intervention, particularly on anxiety, rather than physical health or social functioning.

Table 18: Adjusted means of scores on the GHQ-28 and the HADS – intervention and non-intervention conditions (Houston et al. 1998)				
Sub-scales <sup>1</sup>	Intervention (N=31) Adjusted means	Non-intervention (N=30) Adjusted means	F	p
GHQ anxiety and insomnia	2.11	5.02	15.48	<0.001
GHQ somatic symptoms	6.29	6.33	0.00	0.988
GHQ severe depression	8.24	9.38	0.36	0.549
GHQ social dysfunction	5.78	3.26	0.63	0.430
HADS anxiety	3.15	6.04	19.99	<0.001
HADS depression	3.27	5.26	8.98	0.004

1. Each sub-scale consists of 7 items rated on four-point scale of 0-3. A score between 0-7 indicates that respondents are not troubled by the issues assessed. Scores between 8 and 14 would indicate that respondents are sometimes troubled by the issues assessed. Scores above 14 would indicate that they are definitely troubled. If the means of the non-intervention group are taken to reflect the pre-test scores for both groups, it is clear that in general the respondents scored relatively low on the scales suggesting that they were not substantially troubled by issues assessed. As observed ranges, means and standard deviations at pre- and post-test are not reported in the paper, it is difficult to assess whether some respondents reported substantial problems with anxiety or depression.

Houston et al. suggest that the differential changes found with the GHQ help to rule out the operation of substantial demand characteristics in their study:

'The significant improvement on only one of the three (sic) GHQ subscales, rules out any possibility that post-intervention responses were in any way influenced by the residents' desire to 'please' the researchers. Had any such effect occurred one would expect uniform improvements across all sub-scales.' (1998:331)

The authors conclude that their study indicates that 'a comparatively small amount of intervention (one hour per week) can have a beneficial impact on the psychological wellbeing of older people in residential care' (1998:331). However, they accept that it is 'not clear how persistent the beneficial effect would be if the intervention was continued over a longer period' (1998:331). They also acknowledge that a number of mechanisms may be at work is accounting for the changes:

‘...other aspects of the intervention apart from its humorous content – social interaction, or, singing – might have produced the observed reduction in anxiety and depression’ (1998:331).

In support of humour being the critical factor, however, they note that:

‘... social interaction was occurring in both intervention and non-intervention groups; and the pilot work determined that the most important aspect of the residents’ enjoyment of the sing-a-long was the fact they found the researchers funny and the activity made them laugh. Furthermore, not all residents joined in with the singing, during the activity’ (1998:331)

### **Critical assessment of Houston et al.**

Houston et al. judge the changes observed in their study in terms of statistical significance, but offer no discussion of the substantive or clinical significance of their findings. It would have been of interest, for example, to know whether any of the residents were judged to be ‘clinically’ anxious or depressed prior to the intervention, and whether any changes occurred on an individual level which brought scores below a clinically significant threshold. On a more concrete level, the GHQ measure of anxiety is described as assessing ‘anxiety and insomnia’ and it would have been of interest to know whether any of the residents experienced better sleep as a result of the intervention.

The authors do not report observed ranges, means and standard deviations for the measures used at pre- and post-test and so it is difficult to assess whether some respondents reported substantial problems with anxiety or depression. Nevertheless, some inferences may be drawn from the data reported. Each of the GHQ-28 and HADS sub-scales employed consist of 7 items rated on a four-point scale of 0-3. A score between 0-7 indicates, therefore, that respondents are not troubled by the issues assessed. Scores between 8 and 14 would indicate that respondents are sometimes troubled by the issues assessed, and scores above 14 would indicate that they are definitely troubled. If the adjusted means reported for the non-intervention group are taken to reflect the pre-test scores for both groups, it is clear from Table 18 that the respondents generally scored relatively low on the scales employed (with the possible exception of the GHQ severe depression scale). This suggests that they were not substantially affected by anxiety or depression.

The main point here is not to dispute that ‘statistically significant’ changes were observed, but to question whether the observed changes actually represent a substantial and meaningful improvement in the residents’ ‘psychological wellbeing’. In this respect, it is unfortunate that the paper provides no qualitative feedback from the participants. It would also have been of considerable interest to know whether care staff in the homes noticed beneficial impacts from the intervention which generalised beyond the activity itself.

Finally, Houston et al.’s reflections on the mechanisms involved in the ‘sing-a-long’ activity are of considerable interest. The current review is focused on evaluating the general hypothesis that singing can be beneficial for wellbeing and health, and

perhaps the most significant aspect of the Houston study for this review is their conclusion that the generation of laughter, rather than social interaction and 'singing' was the crucial factor accounting for the effects of their intervention.

#### **4.5.2 Effects of a community choir for older people**

Without doubt, the study reported by Cohen et al. (2006, 2007) ranks as the most important piece of research on singing and health yet published. The investigation of the impact of active participation in singing was part of a larger ambitious investigation aimed at 'measuring the impact of professionally conducted, community-based cultural programs on the general health, mental health and social activities of older adults aged 65 and older.' (2006: 726) The study was conducted at three separate sites and the programmes of activity ranged from 'painting, writing, poetry, jewelry making and material culture, to music in the form of singing in chorales.' (2006:726)

For the 'singing in chorales' study, 166 'healthy ambulatory older adults' were recruited from the same residential areas in Washington DC to participate in a quasi-experimental study. Of these, 90 served as an 'intervention group' and were involved in professionally facilitated singing activities and rehearsals for 30 weeks a year over two years, plus public performances. The remaining 76 served as a 'comparison group' who received no form of intervention except the assessments involved. Participants' health and social activities were measured before the start of the intervention and then one year and two years later. Health measures included a self assessed overall health rating, information on health service utilisation, medication, number of falls and standardised questionnaires assessing morale, depression and loneliness. In addition participants were asked to give detailed information on the 'nature, frequency and duration' of their social activities.

The outcomes of the study appear quite remarkable. The researchers sum up their findings from the first year of the intervention in the following way:

'Results obtained from utilizing established assessment questionnaires and self-reported measures, controlling for any baseline differences, revealed positive findings for the intervention such that the intervention group (*chorale*) reported a higher overall rating of physical health, fewer doctor visits, less medication use, fewer instances of falls, and fewer other health problems than the comparison group. The intervention group also evidenced better morale and less loneliness than the comparison group. In terms of activity level, the comparison group experienced a significant decline in total number of activities, whereas the intervention group reported a trend toward increased activity.' (2006:726)

Findings continued to be positive after two years, and are summed up as follows:

'Results revealed positive intervention effects in relation to physical health, number of doctor visits, medication usage, depression, morale, and activity level.' (2007:5)

Cohen et al. (2007) highlight the important implications of the improvements seen, given the high average age of the participants:

‘Moreover, the actual improvement reported in general health and the sustained level of involvement in overall activities 2 years into the study among subjects with an average age greater than life expectancy, reflects a reduction in risk factors driving the need for long-term care, through continuing involvement in a high-quality participatory art program – in this case, in an ongoing chorale directed by a professional conductor.’ (2007: 20)

As indicated above, Cohen et al. report findings from the chorale study over two years. The 2006 paper reports findings from the first twelve month follow up and the 2007 paper reports findings after two years.

The results reported in the 2006 will be considered first, followed by a comparison between the data reported in the 2006 and 2007 papers and a consideration of the findings on the two-year follow up.

The analytic approaches employed in the two papers are different, and have some features which are questionable. There are also some inconsistencies and puzzling features which come to light when the information and results in the 2006 and 2007 papers are closely compared.

## **Cohen et al. 2006**

### Theoretical perspective

One of the strengths of the Cohen et al. research is the presentation of a clear theoretical framework. This argues that involvement in creative activity among older people is beneficial for wellbeing and health because it helps to generate a ‘sense of control’ and provides opportunities for ‘social engagement’. Previous research on aging is cited as demonstrating positive health outcomes when older adults experience a sense of control in their lives and are involved in activities which provide ‘meaningful social engagement with others.’ Cohen et al. focus on the value of arts activities because they believe ‘their beauty and productivity’ serve to ‘foster sustained involvement’. They explain the significance of this factor as follows:

‘Many general activities do not have this highly engaging and sustaining quality. Apart from the underlying mechanisms of sense of control and social participation, the amount of exposure to these factors is critical in relation to positive health effects, again reflecting the importance of the sustaining factor of an intervention. Analogous to the impact of physical exercise, many people seek involvement for the natural appeal of the art; secondary positive health benefits are an added bonus.’ (2006:728)

The study undertaken is concerned with testing whether engagement in choral singing results in measurable health benefits, however, and does not attempt to directly test the theoretical framework presented as the foundation of the intervention. No data are reported, for example, on the extent to which participants gained a

greater sense of control in their lives and whether this mediated any improvement in their wellbeing or health. A further point of interest in the context of the present review is that Cohen et al. give no specific attention to whether group singing involves distinct mechanisms which carry potential benefits for health. Singing in a group may be sustained because of its 'beauty and productivity' and it may help to develop a sense of control and increase social engagement, but this may be true for a wide range of alternative activities. The lack of attention to the issue of specific mechanisms brought into play by singing is also reflected in the fact that Cohen et al. provide no information on the details of the singing activities engaged in by the intervention group.

### Data reporting and analysis

Table 1 in the 2006 paper reports the means and standard deviations for measures employed at baseline for the intervention and comparison groups. Table 2 reports results for the two groups at 12-month follow-up. The researchers report that at baseline some of the measures showed statistically significant differences between the two groups but some did not. The approach adopted in looking at the 12-month follow up data was to make direct comparisons between the two groups on those measures initially showing no significant differences at baseline. Where differences were apparent initially, analysis of covariance was used to 'control' for baseline differences.

There are a number of problems, however, with the presentation of data, the analytic strategy employed and the use of significance levels in drawing conclusions from the results.

- Firstly, the results are consistently reported as means and standard deviations, when for some of the measures this would seem inappropriate. This is particularly clear for the falls data for the previous 12 months, where one might have expected to see simple frequencies reported. For the intervention group at base line, for example, the mean falls are given as 0.4 with a standard deviation of 0.93. This mean could indicate that 40% of the sample had fallen once during the previous year and the remainder had not fallen at all. However, the standard deviation is nearly one, and if we assume that most of the data would fall within two standard deviations of the mean, the likelihood is that some respondents had fallen at least twice and maybe even three times. Clearly, a more appropriate form of analysis would have been to simply report the numbers of respondents falling or not falling over the 12 months prior to the study, and during the twelve months of the study itself.
- Secondly, the analysis at baseline involves the total sample of people recruited into the study (N=166), whereas at 12-month follow up, data were available for 141 participants, which represents a loss of 25 participants (i.e. a 15% attrition rate). In cases where no significant difference between the groups was apparent at baseline, and a significant difference emerges at follow-up, it is possible that the difference could be attributable in part to attrition and this is not ruled out. It is also difficult to assess whether this issue might also have affected the results obtained from using analysis of covariance.

- Thirdly, the researchers adopt a fairly liberal significance level of 10% in performing their statistical analyses, because of the ‘exploratory nature of this study.’ It is somewhat surprising that they characterise the study in this way given that the investigation is theoretically grounded, fairly well controlled, involves a range of standardised instruments and uses statistical techniques to reject a series of implicit, but clear, null hypotheses. Nevertheless, a case can be made for using a significance level more stringent than the conventional 5% level, in order to reduce the likelihood of type I errors and gain greater confidence in conclusions that the intervention had an effect. In Table 2, nine significant effects are indicated with p values between 0.1 and 0.01, and of these only 2 achieve significance at the 0.01 level (overall health rating and over the counter medications).
- Fourthly, Cohen et al. are not explicit as to whether one-tailed or two-tailed tests are applied when testing mean differences, although the t-values reported for some post-test comparisons suggests that lower one-tailed values were employed. Given the view that the study is ‘exploratory’ however, it would have been more appropriate to consistently employ two-tailed values in assessing significance. For the overall health rating a one-tailed t-test is employed and application of a two-tailed criterion would reduce the p value to 0.05.
- Fifthly, the researchers give no attention to possible ceiling and floor effects in their measures nor to the issue of skewed distributions. The possibility that these issues are a real consideration is particularly clear for the ‘depression’ scale employed – the Geriatric Depression Scale – Short Form. In describing this measure, they report that the data for the entire sample had a range of 0-10, a mean of 1.73 and a standard deviation of 1.97. As is typical for most self-assessment questionnaires, possible scores on this scale are discrete – 0, 1, 2, 3 etc., and a mean of 1.73 indicates that on average participants are scoring themselves between 1 and 2 on an 11-point scale. In other words, on average, the study participants give no indication of being depressed. However, the standard deviation is close to two, and if we assume that a large majority of respondents fall within two standard deviations either side of the mean, this give a range of 8 points. Clearly, respondents cannot score more than one standard deviation below the mean as this is the start of the scale, and so there is likely to be a distinct positive skew in the data with a few respondents scoring at the midpoint on the scale or beyond. From a mental health point of view, respondents scoring relatively highly may be experiencing problems with depression. It would have been relevant to know whether this was the case, and also whether those people who had higher depression scores initially, showed any change in their scores on the depression scale that might have indicated improvement. No such data are reported by Cohen et al. This example also highlights the limitations of relying upon tests of statistical significance to judge the impact of an intervention, when the real interest lies in whether clinically significant changes have taken place on an individual level.

- Sixthly, it is surprising, given the controlled and quantitative nature of this study that Cohen et al. give no attention to the power of their investigation and do not report effect sizes for the various measures employed. Given the use of well-known standardised measures it would have been a simple matter to estimate the size of likely changes (and even clinically significant changes) and the minimum sample size required for a satisfactory level of power.

In addition to these technical issues, Cohen et al. do not address the possibility of bias due to study demand characteristics, particularly among the participants in the chorale intervention. The likelihood of such bias is clear from the account Cohen et al. give of how people were recruited for the chorale arm of the study:

‘The notice for the intervention group differed only in that it sought singers for a chorale; no singing experience was required, and the study’s purpose was to explore the impact of this activity on general health and mental health as well as involvement in overall individual and group activities...’  
(2006: 728)

It is of course essential from an ethical point of view that participants are given information about the nature of a study before they agree to participate. The issue here, however, is the likelihood that most participants would assume, even if it were not explicitly stated by the investigators, that the activity of singing is thought to be beneficial. Inasmuch as participants enjoy and value the activity of singing, they are likely either consciously or unconsciously to adopt a bias in favour of the hypothesis under test. Such a bias may well have affected the way in which they responded to the questionnaires and scales employed and their preparedness to disclose information about health service utilisation, medication and social activities.



## Findings reported in the 2006 paper

Turning now to the results reported by Cohen et al. (2006). The title of the paper indicates three areas of interest – physical health, mental health and social functioning. Detailed attention will be given here to selected results on physical and mental health as these measures are most pertinent to the hypothesis that singing has health benefits.

### Physical health

If a 1% significance level is taken to indicate significant results, there is evidence that participation in singing had a significant impact on two measures of physical health:

- Overall health rating
- Number of over-the-counter medications

It is also of interest to consider the most striking findings from the Cohen et al. study – namely that reported falls were reduced in the intervention group - even though the difference is reported as being significant at the 5% level.

Table 19 reproduces the findings for these measures reported by Cohen et al. in Tables 1 and 2 in their 2006 paper. In addition, an attempt is made to estimate the effect sizes for the changes observed between base-line and follow up for the intervention and control group, and also the effect size based on the results reported for the two groups at follow-up. These figures are approximations and may well be inflated due to the effects of attrition from baseline to follow up.

A number of critical observations can be made on these results, which elaborate on the general criticisms made above.

Overall health ratings were made on a scale of 0-10, and it is clear that for both the comparison and intervention groups the mean ratings were high (between 7 and 8). This suggests a potential ceiling effect and restricted possibilities for seeing a significant increase in ratings. In fact, for the intervention group, the mean ratings effectively remain unchanged, while the comparison group which shows a small reduction in means. The effect size for the comparison group is small however. No significant difference in means is found at baseline, and so Cohen et al. use this to justify the use of a simple t-test on the follow up data and report a significant difference at the 1% level. It is clear, however, that they are employing a one-tailed critical value in this test. Given the claimed 'exploratory' nature of the study a one-tailed test is probably inappropriate and in any event, the use of a two-tailed test would be more cautious. Caution would also dictate using a more stringent probability value for judging significance than the conventional 5% level. If both are applied, the difference found at follow up would not be statistically significant. It is therefore academic to consider the possible effects of attrition and other biases operating in the study. As this is the strongest finding reported using t-tests, it follows that a more cautious approach to statistical inference means that no significant results were found at follow up on the basis of simple t-tests.

Over-the-counter medications data are simple means of such medications reported by participants. The results appear to show that for both intervention and control groups average medication levels increased from pre-test to post-test, with moderate effect sizes. Cohen et al. report that no significant difference was apparent at baseline and their analytical strategy would suggest that a t-test would again be used to compare means at follow up. Somewhat surprisingly, however, an F value is reported indicating the use of analysis of covariance. The estimated effect size for this variable is the largest found in the study, and points to greater use of medication among the comparison group than the intervention group. It is difficult to evaluate this finding statistically as it is not clear whether a one or two-tailed test was applied. There is also no way of ruling out the possibility that the difference reflects the effects of attrition and other biases operating in the study.

Number of falls is reported as means and it was suggested above that this is inappropriate. Nevertheless, Cohen et al. indicate that no significant difference was found at baseline and mean number of falls was reduced in the intervention group and increased in the comparison group. The effect sizes for the two groups, however, are low. At follow up, a t-test is used to compare the groups and a statistically significant difference is reported at the 5% level. Again, however, a one-tailed test appears to have been applied, and the use of a more stringent two-tailed criterion means that the difference at follow up is no longer significant.

### Mental health

Three measures of mental health were employed in the study – morale, depression and loneliness. Table 20 reproduces the findings for these measures reported by Cohen et al. in Tables 1 and 2 in their 2006 paper. As with the physical health measures, estimates are given of effect sizes observed between base-line and follow up for the intervention and control group, and based on the results reported for the two groups at follow-up. These figures are approximations but clearly are very low from pre to post test, and no more than moderate when comparisons are made at follow up. The moderate effect sizes may well be inflated due to attrition from baseline to follow up and the operation of other biases created by the demand characteristics of the study.

The overall picture obtained from the mental health measures is very disappointing as none of the differences reported are significant at even the 5% level. If a more cautious approach is adopted using two-tailed tests at the 1% level, then clearly there is no evidence that the singing intervention had any effect on participants' level of mental and social wellbeing.

A large factor accounting for the lack of measured effects may well be the operation of floor and ceiling effects. Overall, the measures indicate that participants had good morale, were not depressed and were not lonely, and there may well have been little real scope for seeing substantial movement given the instruments employed. Equally, though, there is clearly variation on each of the scales, and it may well be that a few people involved in the study were indeed low in morale, depressed and lonely. Unfortunately, however, Cohen et al. provide no information on the number of such people in their study, and whether the intervention had any positive effects for them.

Table 19: Results from Cohen et al. 2006 on Physical Health					
Sample	Condition	Measure employed Mean (SD)		Statistical analysis	Effect size post to pre Cohen's $d^2$
		Before	After (12-months)		
Elderly community members (N=166 at baseline, N=151 at 12- months follow up)		Overall health rating			
	Intervention	7.88 (1.50) N=90	7.97 (1.58) N=77	Baseline difference between intervention and control groups, not significant.	0.06
	Comparison	7.63 (1.71) N=76	7.25 (1.91) N=64	12-month difference between intervention and control groups, reported as significant, $t = -2.39$ , $p = 0.01^1$	-0.21
			Cohen's $d = 0.41^3$		
		Over-the-counter medications			
	Intervention	2.01 (1.77) N=90	2.61 (2.13) N=77	Baseline difference between intervention and control groups, not significant.	0.31
	Comparison	2.66 (2.18) N=76	4.25 (4.60) N=64	12-month difference between intervention and control groups, reported as significant, $F = 10.02$ , $p < 0.01^5$	0.44
			Cohen's $d = -0.46^6$		
		Number of falls			
	Intervention	0.40 (0.93) N=90	0.23 (0.69) N=77	Baseline difference between intervention and control groups, not significant.	-0.21
	Comparison	0.36 (0.82) N=76	0.55 (1.30) N=64	12-month difference between intervention and control groups, reported as significant, $t = 1.82$ , $p = 0.05$	0.17
			Cohen's $d = -0.31^6$		

#### Table 19 Footnotes

1. For samples over 120, the critical value of  $t$  for the 2-tailed 5% level of significance is 1.96. For the 1% level the value is 2.58. Equivalent values for a 1-tailed test are 1.65 and 2.33. It is clear, therefore, that Cohen et al. employed 1-tailed tests of significance in assessing differences at post-test for 'overall health rating' and 'number of falls'. Given the claimed 'exploratory' nature of the study, and the lack of specific directional hypotheses prior to data collection, the use of 1-tailed tests is at best questionable.

2. Cohen (1998) is not the same Cohen as the author of the article under consideration. Cohen's  $d$  is a measure of effect size in which the difference between two means is divided by the average of their standard deviations. The  $d$  values in this column are calculated on the basis of reported means for the three physical health measures at baseline and after 12-months. Values are negative if means fall and positive if means rise. The values are

approximate as Cohen et al. report data for a sample of 166 participants assessed at baseline, and a reduced sample of 141 at 12-months. The attrition of 25 participants (15% of the original sample) is not commented on and data for the reduced sample at baseline are not reported. Nevertheless, the Cohen values provide some indication of the strength of changes observed, particularly in the intervention group. Clearly, effect sizes are low for both groups across for the overall health rating and falls measures. For the over the counter medications measure, an increase occurs in both the intervention and comparison groups and effect sizes are low/moderate.

3. The Cohen d value is calculated for the 12-month data on overall health rating for the intervention and comparison groups. The baseline difference between the two groups is reported as not significant, and so the Cohen d value provides an indication of the effect of the intervention relative to the comparison. Given the lack of comparable data at baseline and 12-months, however, we cannot rule out the possibility that the difference found at 12-months reflects attrition. The effect size found is moderate and appears to show a positive impact of singing on health rating. In fact, the intervention group shows a very marginal increase in mean for health rating over time, while in the comparison group the health rating mean falls.

4. The Cohen d value is calculated for the 12-month data on over the counter medications for the intervention and comparison groups. The baseline difference between the two groups is reported as non-significant. Nevertheless, analysis of covariance appears to have been used and is reported as showing a significant difference at post-test. Interestingly, medication use appears to rise in both groups. The possibility that the difference found at 12-months reflects attrition cannot be ruled out, and at best the effect size is moderate.

5. It appears that analysis of covariance was employed here to assess whether there was a statistically significant difference at follow up. This is surprising as no significant difference was found at baseline, and Cohen et al. report that where no significant difference was found at baseline, t-test or chi-square was employed to assess differences at follow up

6. The Cohen d value is calculated for the 12-month data on falls for the intervention and comparison groups. The baseline difference between the two groups is reported as not significant. A statistically significant difference at the 5% level is reported at post-test. As noted above, this is based on the use of a one-tailed test. The possibility that the difference found at 12-months reflects attrition cannot be ruled out, and at best the effect size is low/moderate

Table 20: Results from Cohen et al. 2006 on Mental Health					
Sample	Condition	Measure employed Mean (SD)		Statistical analysis	Effect size post to pre Cohen's $d^2$
		Before	After (12-months)		
Elderly community members (N=166 at baseline, N=151 at 12-months follow up)					
		Morale			
	Intervention	14.15 (2.42) N=90	14.08 (2.66) N=77	Baseline difference between intervention and control groups, not significant.	-0.03
	Comparison	13.51 (3.07) N=76	13.06 (3.29) N=64	12-month difference between intervention and control groups, $t$ = -1.92, $p < 0.06^1$	-0.14
			Cohen's $d^3 = 0.34$		
		Depression			
	Intervention	1.39 (1.66) N=90	1.14 (1.84) N=77	At baseline, comparison group more depressed, $t=2.41$ , $p<0.03$	-0.14
	Comparison	2.12 (2.23) N=76	1.84 (1.89) N=64)	At post-test, analysis of covariance revealed no difference between the groups	-0.14
			Cohen's $d = -0.38$		
		Loneliness			
	Intervention	35.11 (8.09) N=90	34.60 (7.86) N=77	At baseline, comparison group more lonely, $t=2.22$ , $p<0.03$	-0.06
	Comparison	38.26 (10.07) N=76	37.02 (10.33) N=64	At post-test, analysis of covariance revealed intervention group 'marginally' less lonely, $F=3.08$ , $p=0.08$	-0.12
		Cohen's $d = -0.26$			

#### Table 20 Footnotes

1. Cohen et al. report differences at the  $p<0.1$  level as significant on account of the 'exploratory nature' of the study.

2. The Cohen  $d$  values in this column are calculated on the basis of reported means for the three mental health measures at baseline and after 12-months. All values are negative as means fall for all measures and each group (lower depression and loneliness, but also lower morale). The values are approximate as Cohen et al. report data for a sample of 166 participants assessed at baseline, and a reduced sample of 141 at 12-months. The attrition of 25 participants (15% of the original sample) is not commented on and data for the reduced sample at baseline are not reported. Nevertheless, the Cohen values provide some indication of the strength of changes observed, particularly in the intervention group. Clearly, effect sizes are low for both groups across all three measures. The

findings for depression and loneliness indicate lower scores after 12-months (a positive change), but the findings for morale also indicate lower scores after 12-months (a negative change)

3. The Cohen d value is calculated for the 12-month data on morale for the intervention and comparison groups. The baseline difference between the two groups is reported as not significant, and so the Cohen d value provides an indication of the effect of the intervention relative to the comparison. Given the lack of comparable data at baseline and 12-months, however, we cannot rule out the possibility that the difference found at 12-months reflects attrition. The effect size found is low/moderate and appears to show a positive impact of singing on morale. In fact, however, both groups show a fall in means for morale over time, with the comparison group showing a larger fall than the intervention group.

4. The Cohen d value is calculated for the 12-month data on depression for the intervention and comparison groups. The baseline difference between the two groups is reported as significant with the comparison group reporting a higher depression score than the intervention group, and analysis of covariance is said to show no significant difference at post-test. The Cohen d value serves only to indicate that the scale of the potential effect of singing is low/moderate at best.

5. The Cohen d value is calculated for the 12-month data on loneliness for the intervention and comparison groups. The baseline difference between the two groups is reported as significant with the comparison group reporting a higher loneliness score than the intervention group. A 'marginally' significant difference is also reported at post-test with analysis of covariance. Interpretation of the Cohen d value is again complicated by the possibility that the difference found at 12-months reflects attrition cannot be ruled out, but at best the effect size is low/moderate.

## **Cohen et al. 2007**

One of the admirable features of the Cohen et al. study is that the investigation continued over two years, with follow up, as we have seen, after 12 months, and then again after two years. This study is thus the first extensive longitudinal investigation of the potential benefits of singing for wellbeing and health undertaken. Unfortunately however, many of the criticisms raised in relation to the 2006 paper continue to beset the presentation and analysis of findings in the 2007 paper, not least the questionable and relatively lax application of statistical procedures for assessing patterns in the data gathered.

### Attrition over two years

Cohen et al. (2007) do acknowledge the factor of attrition in this paper and decide to focus on findings from a core group of 128 participants with a complete set of data at baseline, and both follow up assessments:

‘... this study sample consisted of 128 participants from the Washington, DC area who had data for all three time points. Thus, results for baseline and first follow-up outcomes reported elsewhere (Cohen et al., 2006) are not identical to the first follow-up outcomes reported here, because they do not reflect the smaller sample size utilized here that resulted from normal attrition of the number of participants during the progression of the study.’ (Cohen et al., 2007:11)

Cohen et al. do not consider the reasons for attrition nor do they attempt to assess whether attrition had any relationship to their measures of physical and mental health (e.g. were those who dropped out in poorer health?). This possibility can be explored, however, by comparing baseline data reported in 2006 and 2007. Table 21 gives these comparisons using the measures of physical and mental health considered earlier in discussing the 2006 paper.

Three points should be made about the data in this table before considering the possible effects of attrition.

- Firstly, it is clear that attrition rates were quite high, varying from 24-37% depending upon the measures. Cohen et al. do not explain the variable rates, but the larger Ns for the physical health measures compared with mental health issues, suggest that in addition to attrition, some participants provided physical health data, but did not, for some reason, complete the mental health questionnaires.
- Secondly, the ‘overall health rating’ is scaled differently in the 2006 and 2007 papers. In 2006, it is reported that participants ‘rated their overall health on a scale of 0-10 (0 being worst and 10 being best)’ (2006: 730), while in the 2007 paper it is stated that self-rated ‘overall health’ was assessed on a four-point scale with ‘1=poor, 2=fair, 3=good, 4=excellent’ (2007: 14). The reason for this change of scaling are not given, nor is the 0-10 scale used initially mapped onto the 1-4 point scale used subsequently. Given the change, however, no

judgement can be made about the relationship between attrition and overall health rating.

- Thirdly, the mental health data for the intervention group reported in the 2006 and 2007 papers are identical to the first decimal place. Given the 37% attrition rate, this degree of similarity is implausible and suggests that there are reporting errors in the 2007 paper.

In considering whether attrition in the Cohen et al. study was related to health at baseline, we clearly have to ignore data on overall health rating and the mental health data for the intervention group.

If elderly people with poorer health at baseline were more likely to drop out of the study, this would undoubtedly have introduced bias into the analysis reported in the 2006 paper as noted earlier. Such attrition would lead to a predictable pattern in the data reported in Table 19, namely that baseline means reported in 2007 would be higher for measures of positive health (i.e. Morale) and lower for indicators of poorer health (e.g. medication, depression).

For the two physical health measures considered, there is no clear indication that attrition was associated with poorer health. For over-the-counter medications, the 2007 baseline figure is lower in the comparison group, but higher in the intervention group. For falls, the mean number of falls is lower in the intervention group, but is unchanged for the comparison group. For the mental health measures reported for the comparison group, however, there is a consistent pattern suggesting that attrition may have been a function of poorer mental health as the morale score is higher in the 2007 baseline and the depression and loneliness scores are lower. These differences support the arguments made above that the approach adopted in the 2006 paper in the analysis of 12-month follow up data, may have been subject to bias due to attrition.



Table 21: Baseline data for selected measures reported in Cohen et al. 2006 and 2007

Sample	Physical health		Intervention		Difference	Comparison		Difference
	Mean (SD)	2006 (N=90)	2007 (N=68)	2006 (N=76)		2007 (N=58)		
Elderly community members (N=166 at baseline, N=151 at 12-months follow up, N=128 at two-year follow up)	Overall health rating	7.88 (1.50)	2.4 (0.68)	N/A <sup>1</sup>	7.63 (1.21)	2.3 (0.82)	N/A	
	Over-the-counter medicines <sup>2</sup>	2.01 (1.77)	2.2 (1.9)	Increase	2.66 (2.18)	2.5 (0.82)	Decrease	
	Falls <sup>3</sup>	0.40 (0.93)	0.34 (0.74)	Decrease	0.36 (0.82)	0.36 (0.85)	No change	
	Mental health	Intervention <sup>4</sup>		Comparison <sup>5</sup>				
	Mean (SD)	2006 (N=90)	2007 (N=57)	2006 (N=76)	2007 (N=55)			
Morale	14.15 (2.42)	14.2 (2.6)	No change	13.51 (3.07)	13.9 (2.9)	Increase		
Depression	1.39 (1.66)	1.4 (1.7)	No change	2.12 (2.23)	1.9 (1.9)	Decrease		
Loneliness	35.11 (8.09)	35.1 (8.1)	No change	38.26 (10.07)	37.6 (9.3)	Decrease		

1 The scaling of this variable is different in the 2006 and 2007 papers. As a result it is not possible to make a direct comparison between the means. No explanation is offered by Cohen et al. (2007) for the change in scaling.

2 Attrition due to poorer initial health at pre-test is assumed to result in lower mean levels of medication usage. As pre-test means increase for the intervention group and decrease for the comparison group, there is no clear evidence that attrition is associated with greater medication usage initially.

3 Attrition due to a prior history of falls is assumed to result in lower mean levels of falls. Pre-test means are lower for the intervention group but remain unchanged for the comparison group. There is thus no clear evidence that attrition is associated with more falls at pre-test.

4 The means and standard deviations reported in the 2006 and 2007 papers are identical to one decimal place for all figures apart from the standard deviation value for the morale measure. Such consistency is unlikely given the reported attrition rate of 37%.

5 Attrition due to poorer mental health at pre-test would result in an increase in morale scores and a decrease in depression and loneliness scores. This pattern is observed here and suggests that poorer mental health may have been a factor accounting for attrition over time.

## Differences in data reporting and analysis between the 2006 and 2007 papers

The results for the same set of measures considered in the 2006 paper are again reported in the 2007 paper, but a number of new variables are added:

- The number of operations participants had undertaken
- The average number of total medications and prescription medications
- The number of health problems reported

These data are reported for the two-year period in the 2007 paper. This indicates that data were available at 12-month follow up but were excluded from the 2006 analysis, presumably on the grounds that no significant effects were observed. Such selective reporting of results is generally regarded as inappropriate.

In addition, different analytical strategies are employed in the 2007 paper. For 'continuous' variables, Cohen et al. employ 'repeated-measures ANOVA:

A repeated-measures ANOVA with one between-group factor (group) was used to examine differences over time between the intervention and comparison groups on all continuous variables. The overall F statistic (correcting for sphericity using Greenhouse-Geisser where required) is reported to evaluate the null hypothesis that group means do not differ across the three time points. (2007: 11)

What is not explicitly stated, however, is that the interaction terms for 'group x time' were taken to indicate different trends over time for the intervention and comparison groups.

At least two significant limitations are apparent in the approach adopted by Cohen et al. in report their two year follow-up results. Firstly, as in the 2006 paper, results at the 10% level are accepted as statistically significant. This is a very liberal criterion, and if a more stringent 1% alpha is adopted, only one interaction reported in the 2007 paper emerges as significant – namely, the total number of medicines consumed (both prescription and over-the-counter). Secondly, however, no post-hoc tests are applied where interaction terms appear significant to assess differences between subgroups and to establish where the interaction occurs.

### **Findings reported in the 2007 paper**

In the remainder of this analysis, attention will be focused on the same set of variables as discussed for the 2006 paper. Table 22 reports the means and standard deviations for measures of physical and mental health as reported by Cohen et al. in their tables 2 (p.12) and 3 (p.13). As Cohen et al. report total medications and prescription medications in addition to 'over-the-counter' medications, results for these measures are also given.

Table 22: Selected results from Cohen et al. (2007) for physical and mental health

Measures	Intervention (N=68)			Comparison (N=58)			Statistical analysis (ANOVA) <sup>1</sup>		
	Baseline	T1	T2	Baseline	T1	T2	Main effects	Interaction	
Physical Health									
Overall health	2.4 (0.68)	2.4 (0.75)	3.2 (0.95)	2.3 (0.82)	2.2 (0.72)	2.9 (1.1)	Group: not reported Time: F=4.85, p<.03	Not reported	
Medications	6.1 (3.4)	7.0 (3.5)	7.5 (3.6)	5.7 (3.3)	8.5 (5.4)	8.3 (4.4)	Not reported	F=5.06, p<.01	
Prescription medications	4.0 (2.9)	4.3 (3.2)	4.6 (3.2)	3.4 (2.8)	4.3 (3.1)	4.8 (3.2)	Not reported	F=2.39, p<.10	
Over the counter medications	2.2 (1.9)	2.6 (2.1)	3.0 (2.2)	2.5 (2.0)	4.2 (4.6)	3.5 (3.1)	Not reported	F=3.77, p<.04	
Falls	0.34 (0.74)	0.25 (0.72)	0.40 (0.91)	0.36 (0.85)	0.58 (1.3)	0.46 (1.0)	Not reported	Not reported	
Measures									
Mental Health									
Morale	14.2 (2.6)	14.3 (2.6)	13.9 (2.7)	13.9 (2.9)	13.3 (3.3)	12.8 (3.3)	Group: F=2.50, p=.10 Time: F=4.23, p<.02	Not reported	
Depression	1.4 (1.7)	1.0 (1.8)	1.4 (1.8)	1.9 (1.9)	1.9 (1.9)	2.2 (2.0)	Group: F=4.66, p<.04 Time: F=3.48, p<.04	Not reported	
Loneliness	35.1 (8.1)	33.8 (7.3)	34.3 (9.4)	37.6 (9.3)	36.0 (9.7)	36.2 (9.8)	Group: not reported Time: F=2.94, p<.06	Not reported	

<sup>1</sup> The hypothesis under test is that the intervention is beneficial to health compared with the non-intervention comparison condition. A number of patterns in the data would be consistent with this hypothesis and rejection of the null hypothesis would require a significant interaction term. The principal problems with the analysis undertaken by Cohen et al. (2007) are: high alpha values raising the risk of type I error, and a lack of post-hoc paired comparisons to clarify the nature of the interaction where one appears to exist. The only interaction reported in the 2007 paper which achieves a p value of 0.01 is for total medications. The data show that for the intervention group, medication levels consistently increase over time, and for the comparison group, levels increase at first follow up and then decline slightly. Medication levels are similar at pre-test and then are greater in the comparison group at T1 and T2, but it is unclear whether these differences are significant due to the lack of post-hoc comparisons.

## Physical health

Self-assessed 'overall health' is said to show a significant main effect due to time (means higher after 2 years than at baseline in both groups), but as there is no significant interaction effect, Cohen et al.'s commentary on the data appears to be little more than wishful thinking:

'...we can see that over time, the intervention group reported better health while the comparison group initially reported a decrease in overall health and then a slight rebound that was still lower than the intervention group.'

Data on medication use are reported as showing significant interaction effects, although as noted above, only one interaction is significant at the 1% level, and no post-hoc tests are applied to clarify the nature of the interaction. Cohen et al. again offer optimistic interpretations of their results for medication. Here, for example, is Cohen et al.'s commentary on the data for 'over-the-counter' medicines:

'In terms of over-the-counter (OTC) medication use, it appears that there was a significant increase in OTC usage within the comparison group from baseline to the first follow up with a slight decrease at the second follow up, where there was only a moderate steady increase in OTC usage for the intervention group.' (2007:14)

The reference to a 'significant increase' in the comparison group between baseline and first follow up cannot be justified unless post-hoc group comparisons are made. In addition, while the apparent shifts over the first 12 months might support the general hypothesis that singing is beneficial to health, the shift from year one to year two follow-up actually shows the opposite trend.

Finally, for the falls data, no significant effects are found, and Cohen et al. offer no commentary on these results. The 12-month follow up shows a decline in falls for the intervention group and a rise for the comparison group. At second follow-up, however, the changes are in the opposite direction. The lack of significant effects and the inconsistent pattern of change suggest again that there is no evidence of an effect from singing.

## Mental health

Cohen et al. report main effects due to group and time for the mental health measures, but these findings are irrelevant to the main hypothesis under test. More importantly, no significant interaction effects were found, and thus, there is no hint of support for the notion that choral singing had a positive impact on the measures of mental health employed. The findings as reported are certainly at odds with the following commentary given by Cohen et al. (2007) in the discussion section of their paper:

'... the trend to overall better mental health scores in the intervention versus comparison group is consistent with the better rating of overall health in the intervention group.' (2007:17)

## 4.6 Quantitative studies that use objective physical performance measures

### 4.6.1 Singing and lung function

The activity of singing clearly relies upon breathing (Titze, 2008), and three studies have investigated the hypothesis that singing has a beneficial effect on aspects of 'pulmonary function.' The two earliest studies, Heller et al. (1960) and Gould and Okamura (1973) are at variance with one another. The former study reported no differences on nine measures of lung function in a comparison of 16 professional singers and 21 participants with no 'professional vocal training.' The second study by contrast, compared trained professional singers, students of voice and participants with no vocal training, and found that singers were better able to use their 'total lung capacity.'

'Specifically, it was found that the ratio of the residual lung volume (RV) (the amount of air remaining in the lungs at the end of a total voluntary expiration) to TLC (total lung capacity) was lower in the trained singers than in the students of voice and that the students in turn, had a lower RV/TLC ratio than untrained subjects.' (1973:89)

Both studies have been strongly criticised by Schorr-Lesnicks et al. (1985), however, for methodological weaknesses due to small sample sizes and inappropriate controls. In their study, the groups compared were 34 professional singers from the New York City Opera and Choristers' Union, 48 wind instrumentalists and 31 string or percussion players with professional ensembles or bands.

Comparisons were made on seven measures of pulmonary function:

- Maximum voluntary ventilation
- Forced expiratory volume in one second (FEV<sub>1</sub>)
- Forced vital capacity (FVC)
- Mean forced expiratory flow during the middle half of FVC
- FEV<sub>1</sub> / FVC
- Peak expiratory pressure
- Peak inspiratory pressure

No significant differences were found between groups when compared directly or when controlling for potentially confounding variables such as weight, smoking, and years of performing. As the authors' state:

'We compared string and percussion instrumentalists with wind players and with vocalists with and without holding confounding factors constant, and in no case was there any significant difference between them in any of the tests of pulmonary function or pressure studies.' (1985:203)

Interesting differences did emerge, however from a 'self-administered questionnaire' designed to assess performance and occupational histories, medical history and health attitudes and behaviours. Schorr-Lesnicks et al.'s summary of the differences found between groups is worth quoting in full:

'While wind players thought they were more subject to stress and nervous conditions, and string and percussion instrumentalists believed that they were more susceptible to muscular aches and fatigue, the singers believed that they were more susceptible to illness, particularly upper-respiratory-tract infections. Singers believed that prevention of illness was possible and actively engaged in exercise and generally abstained from tobacco. This heightened awareness of health may have created or fuelled the belief of improved pulmonary function among singers. Based upon the parameters of pulmonary function studies, we could find no basis in fact for this myth.' (1985: 204)

#### **4.6.2 Choir singing and noise-induced hearing loss**

All music involves the production of sound, and singers of course are their own instrument, producing sound with their voice. Given the size of the vocal apparatus, the volume of sound that can be produced is remarkable (Titze, 2002). In the case of group or choral singing, singers are in close proximity to one another and individual singers not only hear the sounds they make themselves (transmitted through their body and through the air), but can hear the sounds produced by voices around them. Undoubtedly, a large choir can produce a large sound, and in the only study of its kind, Steurer et al. (1998) suggest that the level of sound produced during choir singing may be great enough to produce 'noise-induced hearing loss' if exposure is sufficiently regular and prolonged. They suggest in fact, that peak 'noise' levels in excess of 110 decibels can be produced during choir singing, with the great volumes found between frequencies of 100Hz at the lower end and 1KHz at the higher end (G<sub>4</sub> – the G above middle C, has a frequency of 400 Hz). Music of sound levels higher than 100 decibels is said to be 'dangerous for hearing and increases the probability of hearing loss.'

To explore this possibility, Steurer et al. assessed the 'hearing threshold levels' of 62 singers in the Vienna State Opera Chorus, and compared the results with existing norms for the measures employed, and with a 'control' group of 27 noise exposed workers. They provide the following description of the chorus and the levels of their engagement in rehearsal and performance:

'The 30 female singers were on average 49 years old and had been professional singers for a mean of 27 years. The weekly mean singing and performances times were 29 and 15 hours respectively. The 32 male singers had a mean age of 48 and had been performing as professionals for an average of 24 years. The mean weekly singing time was 28 hours with a mean performance time of 15 hours.' (1998:40)

To assess hearing thresholds, 'pure tone audiometry' was employed, and 'absolute air conduction thresholds' were measured at frequencies of 125, 250, 500 and 750 Hz, and 1, 1.5, 2, 3, 4, 6, and 9 KHz, using the 'ascending method' (125 Hz is approximately 2 octaves below middle C).

The findings from this study are highly detailed and complex, but essentially, raised hearing thresholds were found for the chorus singers compared with available norms, across the assessed frequency range.

As an illustration, Table 23 reports data for men aged 40-49 in the sample (N=18) compared with normative data (ISO2029) for men aged 45.

Table 23: Median hearing threshold levels in decibels (Steurer 1998)								
	250Hz	500Hz	1KHz	2KHz	3KHz	4KHz	6KHz	8KHz
Men 40-49	17.3	12.3	9.3	9.4	11.3	21.3	25.0	24.0
Men 45 ISO7029	2.2	2.6	2.9	5.1	8.4	11.7	13.1	16.0
Difference	15.1	9.9	6.4	4.3	2.9	9.6	11.9	8.0
t-values 1% = 2.72	12.85	8.67	6.05	4.62	3.56	4.95	5.91	4.57

It is clear from the table that the hearing thresholds for this group of male singers was raised across the entire frequency range compared with the normative data, but with a more marked effect at the lower frequencies. It should be noted, however, that the authors provide no commentary on the clinical significance of these differences, and whether, for example, they represent a change which would lead to a noticeable handicap for the singers.

In addition, it might be expected that the length of time singers have been involved in professional singing, and their levels of engagement with it, would have a bearing on hearing loss, but surprisingly, the study failed to find evidence of such relationships:

‘In our study there was no significant correlation between hearing threshold, and age, years of professional performance, and weekly singing and performance times, except for a positive correlation between age and the threshold at 2 and 3 KHz and a negative correlation between weekly singing time and the threshold at 2KHz in women only.’ (1998:49)

The authors suggest that the lack of expected relationships may reflect the limited age range in the sample. Also, current levels of singing and performance may not reflect a singer’s history of engagement. Whatever the reasons for the null results, however, the authors rather weakly suggest that:

‘It still seems plausible that age and years of professional performance, if not weekly performance time, do have an impact on hearing threshold,’ (1998:49)

Steurer et al. conclude:

‘...that the choir singers in our study had impaired hearing thresholds regardless of their sex. [and that] Permanent threshold shifts at 250Hz and above are probably noise induced with choir singing as the noise source.’ (1998:49)

Interestingly, the authors go on to suggest that an additional factor may be involved in hearing loss at the lowest frequencies of 125-250Hz:

‘An unproven hypothesis is that singing might lead to increased endolymphatic pressure which thus might cause hearing loss in the low frequency region.’ (1998:49)

They indicate that the idea that ‘singing increases endocranial pressure, will be the subject of future studies’, but unfortunately, no further work on this issue by this group has appeared in print.

#### **4.7 Epidemiological research on determinants of health**

Two large-scale population surveys (Bygren, Konlaan and Johansson, 1996; Konlaan, Bygren, and Johansson, 2000; Johansson, Konlaan and Bygren, 2001) in Sweden and Hyyppä and Mäki (2001a) in Finland) have investigated the possible health benefits of cultural and social participation, including singing in choirs. Neither study, however, found evidence that engaging in choral singing had any significant relationship to the health measures employed, when other substantial determinants of health were statistically controlled.

In relation to this review, it is of value to examine these studies in some detail, for several reasons:

- The studies are very substantial in size and scope and represent the only pieces of research in which the potential significance of choral singing is examined in large representative population samples
- The studies appear in prestigious medical journals (the British Medical Journal and Preventive Medicine) and were conducted by experienced and highly competent research groups (see e.g. Konlaan et al., 2002; Wilkinson, et al., 2007; Hyyppä and Mäki, 2001b; Hyyppä and Mäki, 2003; Hyyppä et al., 2005) – they are therefore likely to carry more weight among medical / public health professionals than would any number of smaller scale studies
- The factor of whether or not people engage in choral singing is examined in the context of a wide range of additional known physical and social determinants of health, and so in principle at least, the research design could show not only that singing is an independent factor linked to health when other influential factors are controlled for, but also the scale of the impact that choral singing has (i.e. how much difference it makes whether or not people sing)
- Given the first three points, the fact that neither study was able to find evidence of an independent influence of choral singing on health is a substantial challenge to the view that choral singing has important health benefits and deserves to be considered within the field of public health as a health promoting activity. Any number of small scale qualitative studies might suggest this – but if large-scale epidemiological studies of the kind that public health professionals take seriously fail to find evidence of such a connection, choral singing will not be seriously considered as an activity worthy of support.

As we have seen many smaller scale, qualitative studies, do suggest that singing has benefits for wellbeing and health, and that this is further supported by a small number



of experimental studies. Consequently, an important issue for this review is to assess whether or not the findings on choral singing from these large-scale Scandinavian studies are convincing. It will be argued that both studies have methodological problems which mean that they do not satisfactorily address the possible significance of choral singing for health, and that their conclusions cannot be relied upon.

Two qualifications should be expressed in relation to this critique:

- Firstly, for both studies, choral singing was included as one social factor among many that were considered to be of potential significance in relation to health. The studies were not centrally concerned with choral singing and the methodological issues raised when this factor is focused on, is not meant to question the general validity of the studies given their primary purposes.
- Secondly, the authors of this review are very aware that they may be accused of 'special pleading' given their interest in establishing an evidence base to support the notion that choral singing has benefits for wellbeing and health. The attention being given to these studies is a mark of the seriousness with which they are regarded within the entire corpus of research identified for this review. They fall clearly within the domain of public health research, and so would carry weight among public health practitioners and commissioners who are required to consider the 'evidence-base' to support the funding of initiatives to meet current health agendas and targets.

#### **4.7.1 Cultural participation and health in Sweden**

In the wider field of 'culture and health', the study by Bygren, Konlaan and Johansson (1996) is often quoted as supporting the case that engagement in cultural activities is beneficial for wellbeing and health. The aim of the study was to investigate the possible influence of attendance at cultural events, reading books or periodicals, making music or singing in a choir as determinants for survival.

The design was a large-scale population survey in which a total of 12,675 people aged 16-74 years from across Sweden were interviewed in 1982/3 about their leisure activities and 6301 men and 6374 women were followed up until to the end of 1991 to determine 'survival rates' (i.e. the proportion of participants still alive). Konlaan, Bygren and Johansson (2000) subsequently followed up the same cohort until the end of 1996. In addition, Johansson, Konlaan and Bygren (2001) report a related study in which 3793 participants in the original survey were re-interviewed in 1990/1 and data gathered on 'self-reported health.' There are differences in the details of the analysis undertaken in these three papers, but as the findings are essentially comparable with respect to making music and choral singing in showing no links to survival or self-reported health. For this reason, we will limit detailed consideration to the 1996 paper.

## **Bygren et al.1996**

In the baseline survey data were gathered on long-term illnesses, smoking and physical activity and on involvement in leisure and cultural activities. Three scales were constructed to measure involvement in cultural activities. It is important to understand the nature of these indices, so Bygren et al.'s account is reproduced here in full:

The cultural events attendance index was constructed from attending a cinema, theatre, concert and live music, museum, art exhibition, sermon, or sports events as a spectator. The number of visits at each of these seven kinds of events was classified into four classes, with more than 20 visits per year giving four points. These scores were used to classify the cohorts into two low use groups (<9 points and 9-12 points). The most active group scored more than 12 points, making at least 80 visits per year, and constituted the reference group. The second independent variable was constructed out of the questions about reading books or periodicals. Two classes of low active readers were contrasted against a reference group of those reading books or periodicals at least once a week. The third independent variable index was constructed out of questions on how often the respondent played music or sang in a choir, and they were dichotomised into those doing either at least "now and then" and those doing it less often than "now and then." (1996: 1579)

On follow up, 533 men and 314 women had died (i.e. between 1982/3 and 1991) and survival rates were examined in relation to the three indices of cultural involvement with eight potential 'confounders' controlled for: age, sex, weekly contacts with close friends, education, income, long term disease, smoking, and physical exercise.

Only the first index measuring level of attendance at cultural events at baseline emerged as a predictor of survival rates 9-10 years later. The key finding from the study is represented in figure 1. This shows that people in the sample who 'often' attended a range of cultural events were more likely to survive over the period of the study, than were those who initially 'rarely' attended such events. At the start of the study, the survival estimate is 1.0 indicating that all participants are alive, whereas at the conclusion of the study, approximately 2% of those 'often' attending cultural events had died, compared with approximately 4% of those 'rarely' attending such events. The latter group appear to be twice as likely to die as the former group.

The authors attempt to provide a theoretical interpretation of their findings, but their suggestions are rather non-specific and vague. Cultural activities, they suggest, may be inspirational for dealing with problems by 'enhancing people's reflection on their life situation and enabling them to prepare for coming events in their mind.' In addition, cultural activities may provide 'direct vicarious emotional arousal without damage or side effects in real life, in the same way that 'verbal expression of traumatic experiences through writing or talking improves physical health, enhances immune function and is associated with fewer medical visits.'

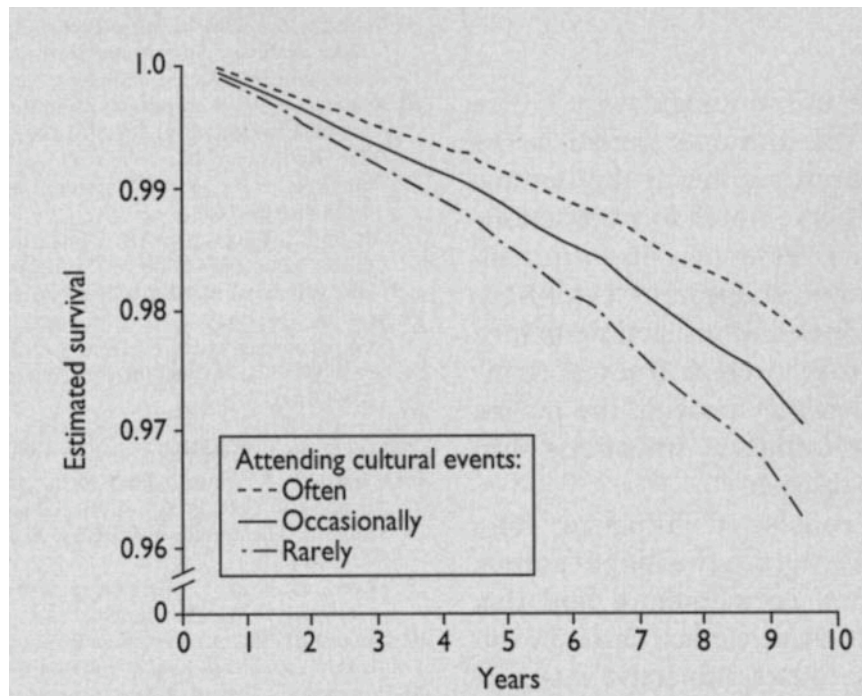


Figure 1: Estimated survival of people attending cultural events often, occasionally, or rarely. Eight confounders were controlled for (age, sex, weekly contacts with close friends, education, income, long term disease, smoking, and physical exercise) (from Bygren et al., 1996)

### Limitations of the study

The authors acknowledge that their 'analysis of the influence of cultural stimulation ... took advantage of a standard of living study that was not originally designed for such a purpose. Therefore the frequencies of visits or activities could be seen as only crude measures of such stimulation because the emotional responses to or perceptions of the visits are unknown.'

Interestingly, they state - 'We could have anticipated a more distinct influence in the groups that were more active making music or singing in a choir than in the more passive activities; perhaps this grouping was not homogeneous enough.'

Undoubtedly, the crudeness of the measures employed is the principal weakness of this study and this deserves further commentary.

- No justification is provided for the way in which the indices of cultural engagement were constructed and no evidence is offered regarding the reliability or validity of these indices.
- The cultural attendance index is remarkably diverse in its content – including not only attendance at arts events, but also attending sports events and church services. It is likely, therefore, that the group of individuals in the high cultural attendance group were made up of a diverse sample of people who were either regular supporters of a sports team, or a committed church goer, or an enthusiastic supporter of arts and music events.

- The index on active involvement in music was a simple dichotomous variable assessing whether participants played instruments or engaged in choral singing 'at least now and then' as opposed to 'less often'. Just over 20% are described as 'making music or singing in a choir' (i.e. 2560 people) and data are not separated for those involved in singing, so no specific conclusions can be drawn about survival rates for singers. Making music 'at least now and then' would cover a large spectrum of frequency and intensity of activity.
- In addition, the measures of cultural involvement relate only to the time of the initial survey, and no data were gathered on the extent of continued involvement over the intervening period before the examination of survival rate.
- The focus on survival was also crude and no information is provided on the causes of death. People would have died from the entire range of possible causes of death - some of which would have no conceivable connection with cultural participation.
- In this regard, it should be noted that the majority of people dying, were in the oldest age group of 65-74 yrs. It would surely have been of value to focus on this age range in more detail to assess the impact of cultural participation for them specifically.

Given the focus in this review on the potential value of singing for health, it is important to question the credibility of the approach adopted in the current study. We can do this by considering more closely the data Bygren et al. report in Table 1 of their paper. Here, the death rate among those involved in music making and singing is given as 4.1%, whereas the death rate among those not involved in making music and singing is 7.3% (see Table 24 below). On the face of it, this difference could suggest a survival benefit associated with involvement in music-making and singing. The authors are quite right, however, to seek to take account of possible 'confounding' variables that could give rise to this simple association.

Unfortunately, no additional information is given about the social and health-related characteristics of the music making and non-music making groups, but it is to be expected that there would be larger proportion of women in the musically active group, compared with the less active group. Since, mortality rates were higher among men and than women (8.5% of men died vs. 4.9% of women), this would clearly account for some of the difference. With respect to age, it is likely that older people would be more represented in the music making group compared with the less active group, but older people were more likely to die than younger. Interestingly, when both sex and age are controlled for, in the first proportional hazards model reported in Table 2 of the paper, the relative risks for mortality associated with involvement with music making appear to be significantly lower (RR = 0.77, 95% confidence interval = 0.63-0.94). The authors go on in a second analysis to take account of further 'confounding' variables. These include education, income, extent of social network, long-term illness, smoking and exercise, but also appear to include the other two cultural participation indices (attending cultural events and reading books or periodicals). For the second model, the lower relative risks for

mortality associated with involvement in music making appear to be non-significant (RR = 0.89, 95% confidence interval = 0.72-1.10)

Table 24: Selected data reported by Bygren et al.1996			
Variable	N	No. of deaths	%
Sex			
Female	6530	314	4.8
Male	6145	533	8.7
Age			
16-44	6530	77	1.2
45-64	3866	306	7.9
65-74	1711	464	27.1
Making music			
Sometimes	2560	104	4.1
Rarely	10115	743	7.3

The account in the paragraph above demonstrates how marked the contrast is between the paucity of the data gathered and the weightiness of the analysis undertaken. It surely strains plausibility to imagine that such a crude measure of 'music making or singing in a choir' assessed in 1982/3 could emerge as a significant, independent predictor of the likelihood of dying over the following 9-10 years, when a range of variables known to be powerful determinants of health and longevity are controlled for. So implausible, in fact, that the failure to find such an association cannot be damaging to the basic hypothesis explored in this review, that singing is potentially beneficial for wellbeing and health.

#### 4.7.2 Leisure activities, social capital and health in Finland

##### Hyypä and Mäki (2001a)

Hyypä and Mäki and their colleagues have undertaken a number of significant studies exploring the value of social capital and leisure activities for health. The findings of these studies largely support the work of the Swedish group, and work undertaken in the United States on the value for health of social and productive activity in addition to physical exercise (e.g. Glass et al., 1999).

In the study under consideration here, the focus was on exploring the possible role of social capital in accounting for the fact that Swedish speaking Finns have better health than Finnish speaking Finns, in the same bilingual region of Finland.

Randomly selected samples of 1,000 adult Swedish speakers and Finnish speakers were invited to participate in a self-completion questionnaire survey (overall response rate 64%) A wide range of socio-demographic, health, health behaviour and social capital data were gathered, and simple comparisons between the Finnish and Swedish speaking samples revealed significant differences on socio-demographic and social capital variables.

The central issue explored in this paper is the extent to which health, socio-demographic and social capital measures predicted 'self-reported health'. Self-reported health was assessed on a five-point scale, but for the purpose of the

analysis this was dichotomised between the highest rating of 'good' vs. the remaining options on the scale – 'almost good / fair / poor / bad'. Self-ratings of health as 'good' were given by a majority of the sample. The analysis was conducted separately for men and women within the two language communities, and the central findings are reported in Table 3 in the paper, for which the authors offer the following commentary:

'The final model (Table 3), in which data were disaggregated by gender and language, shows that after controlling for language, age, BMI, household income, smoking, migration, urban residence and long-term diseases, the number of auxiliary (willing to help) friends and participation in a religious association, were significantly (...) and independently related to self-rated health, whereas singing in a choir, membership of any voluntary association and participation in community events did not predict self-rated health.' (p.152) Further on in the paper they say that 'Singing in a choir associated significantly with the Swedish language in women, although it does not predict self-rated health.' (2001:153)

In terms of the hypothesis explored through this review, that group singing has potential benefits for wellbeing and health of singing, the conclusion drawn is disappointing. However, a number of points need to be considered in critically assessing the validity of this conclusion.

### **Limitations of the Study**

- No information is given on the extent of respondents' involvement in choral singing. Hyypä and Mäki do state that 'singing in a choir is typical among Swedish speaking women', but no indication is given of the percentages of choral singers in the language group / sex samples.
- The choral singing variable is dichotomous (singing in a choir vs. not) as is the health variable (health 'good' vs. not), so no information is given on the nature of involvement in choral singing (e.g. length of time involved, frequency of participation etc.). 'Singing' is an occasional activity in the lives of even the most committed choral singer, but membership of a choir may represent a significant component in the social lives of at least some individual singers.
- No indication is given of whether a simple relationship was observed between choral singing and self-reported health prior to 'controlling for' a wide range of socio-demographic and health variables, which have known implications for self-rated health. As noted in relation to the Bygren et al. paper, factors such as income and BMI has such powerful associations with health, that it would be remarkable indeed if a crude measure of involvement in singing 'survived' as an independent predictor of self-reported health once they were taken into account.
- Even if a relationship were observed, it is clear given the cross-sectional design of the study that any positive association between participation and better self-rated health could not be given a directional interpretation, since participation could be as much a function of health as the reverse. This is an

issue which is not solved by the use of sophisticated statistical techniques to control for other variables.

In personal correspondence, Prof. Hyypä kindly provided the data on percentage involvement in choral singing in the four language / sex groups. These are reported in Table 25 along with the percentages within each group reporting 'good' health.

The level of engagement in choral singing is actually very low, and for Swedish women, the figures are at odds with Hyypä and Mäki's statement in the paper that choral singing is 'typical' in this group. Among men, the levels of engagement are so low that it would be inappropriate to test for even a simple association between singing and self-reported health. For the women, the best case scenario would be that 100% of singers reported 'good' health. If the percentage of 'non-singers' in 'good' health is adjusted accordingly for both samples of women (57% for Swedish women and 64% for Finnish), a statistically significant association emerges, but it is highly unlikely that such a strong pattern would have emerged in the data. If it is assumed that 80% of female singers reported 'good' health, a significant association emerges only for Swedish women (with a 20% differential between singers and non-singers). If the figure is set at 70% for singers, with a differential of 9%, no significant association emerges even for the Swedish women.

Table 25: Choir membership in the national/sex sub-samples in Hyypä and Mäki (2001) (personal communication)				
	Finnish men	Finnish women	Swedish men	Swedish women
N sub-sample	284	374	271	355
% in choirs	1	5	3	11
N in choirs	3	19	8	39
% in 'good' health	65	66	66	62
N in 'good' health	185	247	179	220

Prof. Hyypä did not provide details of the actual differential in reported 'good' health between singers and non-singers, so this is not known, but it is clear, even for Swedish women, that an implausibly high differential of 20% would be required to achieve a statistically significant simple association between singing in a choir and self-reported health. It is fair to conclude, therefore, that the study was under-powered with respect to detecting such an association, and it is certainly inconceivable that such an association would emerge 'unscathed' following statistical control for a range of socio-demographic and health-related variables.

## **5. DISCUSSION**

### **5.1 Identifying sources for this review**

For the purposes of this review, a search was undertaken of ten bibliographic sources using a carefully selected set of search terms organised into three groups: health, music therapy and singing. The strategy adopted was to start with Medline, Embase, Cinahl and Psycinfo as the sources most likely to identify relevant published research reports. A total of 233 papers identified from these sources were categorised on the basis of the coding scheme given in Appendix 2. It was decided at this stage that clinical studies in which singing served as a form of therapeutic intervention for some specific health issue were best considered in a second review, and 16 non-clinical studies were judged to be interest. Further searching of ERIC, ASSIA, RILM, Cochrane Library, Music Therapy World and Google Scholar, produced in excess of 800 references, but only 4 further papers of interest were identified.

A further search strategy involved examining the references given in each of these papers, and this identified a further 9 sources of interest. An attempt was made to contact the authors of papers identified to locate further published material, giving a further 4 sources. A range of additional, often serendipitous routes, served to identify a further 20 papers of potential interest, giving a final total of 54 sources (Table 1).

These papers were then read carefully for content and quality independently by two members of the research team, and a process of categorisation and pruning led to 36 papers identified for systematic mapping and review (Table 2). Studies were grouped in terms of qualitative studies, studies using specifically designed questionnaires, quantitative studies using previously validated scales, studies making objective measures of physiological changes associated with singing (e.g. hormonal changes), studies employing well established measures of 'mental and physical health', research assessing physical performance (i.e. hearing and lung function) of singers and finally large scale epidemiological studies in which choral singing was included as a potential determinant of health.

### **5.2 An initial overview of the corpus of studies**

At this stage, it was very apparent that research interest in singing, wellbeing and health is relatively recent. The earliest paper identified was published in 1960, and most appeared in the late 1990s and beyond. Also, the papers identified were highly variable in terms of problem addressed, the character of singing investigated, participants involved and methods of data gathering and analysis. Many of the qualitative studies were presented as exploratory in character and involved very small samples. The more quantitative studies involving quasi-experimental designs were also small in scale. In addition, the quality of reporting varied with some reports lacking essential details.

Citation mapping within the corpus of sources (Table 3) showed little evidence of being progressive in the sense that researchers are aware of and build upon previously published research. Some of the most recently published studies, for



example, make little reference to previous studies, even where they are clearly relevant. The most obvious examples here are Louhivuori et al. (2005) and Cohen et al. (2006).

Consideration of the character of singing investigated in the studies under review also shows a very wide diversity of focus, with some studies concerned with group or choral singing and others with individual singers; some studies with amateur or professional singers or both, and some with existing singing groups of long standing or with groups set up specifically for research purposes (Table 4). Interestingly, studies varied considerably in the level of detail provided of the singing investigated, with some reports giving little or no account and others offering quite detailed descriptions. The examples given in Table 5 demonstrate vividly that although the corpus of studies are all concerned with 'singing', the actual activities investigated were remarkably diverse.

As the starting point for this review was the interest of the Sidney De Haan Research Centre in the potential value of choral singing for wellbeing and health, studies of group singing were considered of particular relevance. Table 6 identifies 20 reports which have a focus on group singing – whether they were existing amateur choirs or choral societies, professional chorus or choirs or groups specifically established for research purposes. An interesting aspect of these studies is that while a few focus on single sex groups (Ashley, 2002; Bailey and Davidson, 2002b; Silber, 2005) collectively female participants outnumber males by approximately 3:1, and in some studies the imbalance is even greater (Table 7). Interestingly, only one study specifically considered the factor of 'gender' in relation to its findings (Cliff and Hancox, 2001) and reports a significant sex difference in perceived benefits associated with singing. It is clear, therefore, that research in this field is affected by a considerable bias towards females, which should be taken into account in considering any of the findings highlighted in this review.

### **5.3 Benefits associated with singing**

With respect to the possible benefits associated with singing, the studies reviewed fall into two main groups.

Firstly, there are those studies which start from an assumption that singing has value, but which adopt an essentially exploratory, qualitative approach and gather information by asking singers to give accounts of their experience of singing and its personal value for them, through interviews, group discussion and written answers on questionnaires (e.g. Cliff and Hancox, 2001 (first study); Bailey and Davidson, 2002a, 2005). In a few studies, such data is placed in the context of broader ethnographic observation and/or a working knowledge of the participants in the study (e.g. Ashley, 2002; Silber, 2005). Closely associated with these studies are those which gather essentially descriptive data through structured questionnaires (e.g. Hills and Argyle, 1998a,b; Beck et al., 2000; Cliff and Hancox, 2001 (second survey); Hilman, 2002; Bailey and Davidson, 2003b).

Secondly, there are those studies which adopt a hypothesis-testing approach based on the use of previously validated measures of mood, wellbeing or health (e.g. Unwin et al., 2002; Kreutz et al., 2004; Cohen et al., 2006, 2007); objective measures of

biological or physiological functioning assumed to be associated with subjective experiences of wellbeing (e.g Beck et al, 2000; Kuhn, 2002; Grape et al., 2003), or with objective indicators of some aspect of bodily function assumed to be impacted by the activity of singing (e.g. hearing or lung function).

In addition two large-scale epidemiological surveys in Sweden and Finland have included singing in choirs as a possible health determinant (Bygren, Konlaan and Johansson, 1995; Konlaan, Bygren and Johansson, 2000; Johansson, Konlaan and Bygren, 2001; Hyyppä and Mäki, 2001).

### **5.3.1 Benefits identified by qualitative and questionnaire studies**

In relation to the first set of studies, a number of recurrent themes can be identified in the qualitative and descriptive data reported. Namely, that the experience of singing, and particularly singing in a group, has the following perceived benefits:

- Physical relaxation and release of physical tension
- Emotional release and reduction of feelings of stress
- A sense of happiness, positive mood, joy, elation and feeling high
- A sense of greater personal, emotional and physical wellbeing
- An increased sense of arousal and energy
- Stimulation of cognitive capacities – attention, concentration, memory, learning
- A sense of being absorbed by an activity which draws on multiple capacities of the body and the mind
- A sense of collective bonding through coordinated activity following the same pulse
- The potential for personal contact with others who are like-minded and the development of personal supportive friendships and constructive collaborative relationships
- A sense of contributing to a product which is greater than the sum of its parts
- A sense of personal transcendence beyond mundane and everyday realities, being put in touch with a sense of beauty and something beyond words, which is moving or good for the soul
- An increased sense of self-confidence and self-esteem
- A sense of therapeutic benefit in relation to long-standing psychological and social problems (e.g. depression, a history of abuse, problems with drugs and alcohol, social disadvantage)
- A sense of contributing to the wider community through public performance
- A sense of exercising systems of the body through the physical exertion involved in singing – especially the lungs.
- A sense of disciplining the skeletal-muscular system through the adoption of good posture
- Being engaged in a valued, meaningful, worthwhile activity that gives a sense of purpose and motivation

In addition to identifying a wide range of possible benefits, attempts have also been made to organise such a list conceptually and empirically. The most significant conceptual work in this regard has been undertaken by Bailey and Davidson (2002a, 2005), who present a convincing model of ‘the positive effects of participation in

group singing' based on studies of members of two choirs for disadvantaged adults and also middle class choristers. This suggests four principal groups of benefits: clinical-type benefits; social benefits; benefits associated with public performance, and cognitive stimulation. They argue that the therapeutic benefits were widely reported across the samples investigated, but others varied to some extent according to social circumstances.

Interestingly, this model overlaps to some extent with the results reported by Clift and Hancox (2001) from their second survey. Factor analysis of data from a structured questionnaire produced six factors, which they label as: wellbeing and relaxation; breathing and posture; social benefits; spiritual benefits, emotional benefits and heart and immune system benefits. The first factor (together with the fourth and fifth) corresponds closely with the first and fourth component of the Bailey and Davidson model, and the third social factor in Clift and Hancox clearly links with the second social component in Bailey and Davidson. What Bailey and Davidson do not include is the sense of physical health benefits associated with singing, which are represented by Clift and Hancox's second and sixth factors.

Similar correspondence can be seen with the results reported by Hills and Argyle, especially in their 1998a paper. They identify five factors from their musical experience scale, which they label as: Wellbeing, Mystical experience, Social, Entertainment and Intellectual/musical. Clearly, the first factor corresponds with the first component of the Bailey and Davidson model and the first factor in Clift and Hancox. Factors 2 and 3 in Hills and Argyle correspond with similar factors in Clift and Hancox. Their factor 5 also clearly aligns with the component of cognitive stimulation suggested by Bailey and Davidson.

### **5.3.2 Benefits examined in quantitative hypothesis-testing studies**

#### **Singing and mood**

A number of studies have explored the hypothesis that singing generates positive feelings through the use of previously standardised questionnaire designed to assess 'mood' (Rider et al., 1991; Valentine and Evans, 2001; Unwin et al., 2002; Kreutz et al., 2004) (see Tables 13 and 14). In each of these studies, the questionnaires were used to assess 'mood' before a period of singing and then immediately afterwards. Despite the fact that hypothesis under test is essentially the same in each case, the studies are difficult if not impossible to compare as they differ not only in the mood questionnaire employed, but also the characteristics of the singing and participants involved. In addition, methodological difficulties, the possibility of significant demand characteristics and the conceptual problem of linking short-term and slight changes in mood to broader notions of wellbeing and health all raise serious questions over the scientific value of the studies.

The best designed and most relevant study is clearly that reported by Kreutz et al. (2004), and their results do appear to show a significant effect of singing on mood compared with a listening to singing condition. However, close scrutiny of their findings shows that this change is small, when compared with a significant and large shift in negative mood state, when singers are asked to listen to a recording of choral singing rather than sing themselves!

## **Singing and biological / physiological variables**

An obvious attraction of studies which attempt to assess the impact of singing on biological or physiological variables is that they offer the promise of 'objective' knowledge (see Tables 15, 16 and 17). In addition, they may contribute to the development of a theoretical model of the underlying processes through which the activity of singing generates positive feelings and also influences the physical functioning and health of the body. There is also a sense that such objective measures are less likely to be influenced by study 'demand characteristics', although decades of research demonstrating the reality of placebo effects in clinical pharmacological trials, clearly contradicts any assumption that physiological parameters are immune to psychological or social influence!

A clear area of interest pursued in a number of studies is the connection between singing and 'stress'. This actually has two aspects to it. The first is that singing itself can be experienced as 'stressful', particularly when the material being sung is difficult to master, and there are the additional pressures of public performance. Given that it is well known that the experience of 'stress' is mediated by the sympathetic nervous system and the endocrine system, it follows that increases are likely to be found in key hormones involved in the stress response when singers engage in performance. Interestingly, however, none of the studies surveyed which explore physiological parameters in relation to singing, consider the distinction between 'eustress' and 'distress' – in other words, between 'positive' stress, which is important in motivating performance to a high standard, and 'negative' stress which arises when the demands of a task or situation are perceived to be greater than coping resources. The second stance that can be taken in thinking about the relationship between singing and stress, is that singing insofar as it is relaxing and calming, can actually counteract pre-existing feelings of 'being stressed.' Given this position, one might predict that singing would be associated with a lowering of stress hormones.

Both of these positions are illustrated by the study by Beck et al. (2000) in which cortisol levels were assayed in choral singers before and after rehearsals and performance. For rehearsals, cortisol levels reduced, whereas under performance conditions, cortisol levels increased. However, Kreutz et al. (2004) found no changes in cortisol associated with a choral rehearsal, Beck et al. (2006) found no changes associated with individual singing lessons or performances, and Grape et al. (2003) report that cortisol increased for men after a singing lesson, but decreased for women. Overall, therefore, the picture for cortisol changes is somewhat complex, and general conclusions are difficult to draw.

The picture for salivary immunoglobulin A appears to be a little clearer (see Table 16). Five studies have assayed sIgA before and after singing, and the findings suggest that group singing is associated with increased secretion of this anti-body in saliva. As sIgA is described as being part of the immune system which offers 'frontline' defence in mucosal membranes against infection, these findings appear to have some relevance for the health of individual singers. Such a conclusion would be premature, however, given the limitations of the studies. In particular, there is no evidence that such changes measured immediately after singing are sustained, and no evidence that they have any clinical significance. The causal mechanisms involved in generating the changes are also obscure, and it may well be that

increased levels of immunoglobulin are caused not by singing per se, but by the heightened level of 'stress' involved in singing or the fact of being in close proximity of other people who are breathing energetically, and exhaling potentially infectious organisms.

### **Singing and measures of mental and physical health**

Only two quasi-experimental studies have been reported in which the effects of singing on standardised and objective measures of wellbeing and health are assessed (Houston et al., 1998; Cohen, 2006, 2007). In the former study, a four-week 'old time sing-a-long' programme was set up in 3 residential care homes, with 3 homes acting as non-intervention controls. Participants completed two well-known health questionnaires before and after the intervention – the Hospital Anxiety and Depression Scale and the General Health Questionnaire-28 (see Table 18). It should be noted, however, that while singing was a central part of the activity established in the project, the focus of the researchers' attention was on the potential of singing funny songs in generating laughter, which they saw as the active ingredient in the intervention. In the Cohen et al. study, a fairly large community chorale was established for retired people living independently, with a similar number non-participating community members acting as a comparison group. A range of measures were administered before the start of the intervention, and then again after one and two years (see Tables 19-22).

In both cases, positive findings are reported which appear to suggest that involvement in group singing activity had a positive impact on measures of both mental and physical health. These studies are among the best pieces of research available to date which have attempted to directly assess the value of singing for health, and for this reason they were subject to rigorous critical scrutiny in this review.

Both studies have major limitations.

The first is the fact that they employ non-equivalent group designs – and although both research teams present data to suggest that the intervention and control groups were comparable at baseline on the measures employed, it nevertheless remains the case that the differential operation of additional factors that could result in differences emerging between the groups at post-test cannot be ruled out. Only randomised studies can do this.

Secondly, there is no indication that the studies were 'single-blinded' – in other words, that the researchers assessing participants at pre and post-test, and the researchers analysing the data gathered, were blind as to the groups they belonged to.

Thirdly, the participants themselves were clearly aware of whether they were participating in the intervention or were in the control arm of the study, and as such, they might well have developed their own assumptions regarding the purpose of the research, and the outcomes anticipated by the researchers. Under these circumstances, the study results are likely to have been affected by study 'demand

characteristics.’ To their credit, Houston et al. (1998) do recognise this possibility, but Cohen et al. (2006, 2007) do not.

Fourthly, both studies are fairly small in size and given the small to medium effect sizes involved for the measures employed, it is clearly that both are substantially underpowered. Houston et al. (1998) do give some attention to power and effect sizes in a footnote, but give no indication that considerations of study power guided the initial design of the study and decisions about sample size. Similarly, there is no indication that consideration of likely effect sizes and power guided decisions about sample sizes in the Cohen et al. study.

Finally, both studies rely among tests of statistical significance in drawing conclusions about the possible beneficial effects of their interventions on health, but neither seriously addresses the question of whether the changes observed are substantively or clinically significant. This issue comes to the fore particularly clearly with respect to measures of ‘depression’ employed in both studies, as it not established that any of the participants was affected by real feelings of depression prior to the intervention, nor is demonstrated that the mean changes observed in relation to the intervention represent a clinically significant shift nor that any individuals within the intervention group experienced a meaningful move away from depression.

### **Singing, lung function and hearing**

Singing as an activity involves the singer using their own bodies as an instrument. In common with all other musical instruments, this involves three components – a sound source, a sound resonator and a sound radiator (Titze, 2008). It is surprising, therefore, that so few studies have explored the implications of this fundamental fact about singing, and assessed the extent to which the exercise of the basic structures and mechanisms intrinsically involved in singing might lead either to improvements in the fitness and so ‘health’ of the systems involved or to possible damage.

The power for the sound source (the larynx) clearly comes from the lungs – and it follows that one might expect that active involvement in singing may have effects on the structure and function of the lungs and the larynx . It is also obvious that singers produces sound, and that singers in a group can produce quite a substantial volume of sound. There is thus a possibility that regular exposure to high levels of sound produced by a choral singing group may have affects on the structure and functioning of the auditory system. As with the issue of physiological changes associated with singing, research on these questions is attractive given that objective measures are possible. Given also that any structural or functional changes would have to be a result of factors operating over a substantial period of time, it is difficult to see how their assessment could be affected by psychological or social demand characteristics in any study.

Both issues have received research attention, but it is fairly remarkable that so little research has been undertaken. The three earliest studies identified in this review explored the idea that professional singing would lead to a measurable improvement of lung function. The best of these (Schorr-Lesnicks, et al., 1985) appears to have dismissed the idea as ‘a myth’ so conclusively that no further research has been

undertaken on the subject has been undertaken since. Given the lack of evidence for professional singers, it might seem to be unlikely that amateur singing, whether on an individual or group basis would have much impact.

Similarly, only one early study has investigated whether being part of a professional chorus might affect hearing. Steurer, et al.(1998) demonstrate hearing damage, in the sense of higher auditory thresholds, across a wide frequency range (see Table 23). There are some puzzling features of their study, such as the failure to find any relationship between length of career and time devoted to singing – and hearing thresholds, which might seem to call for further investigation. It is also an open question whether such effects would be detected among amateur choral singers.

Both the Schorr-Lesnack, et al. and Steurer et al. studies clearly give pause for thought in relation to the general hypothesis guiding this review – that participation in group singing has benefits for wellbeing and health. In the former case, effects which might be expected - and which as Clift and Hancox (2001) found, were believed to be true among choral singers themselves – are not in fact supported by the evidence. And in the second case, the evidence of raised hearing thresholds raises the possibility that group singing is not necessarily entirely beneficial, but may in fact lead to bodily damage.

### **Epidemiological research with singing as a health determinant**

Two large-scale Scandinavian studies of population health were also considered relevant to the current review (Bygren et al., 1996; Hyyppä and Mäki, 2001). Both explore the idea that social, cultural and leisure activities have a significant impact on population health over and above a range of very well known and powerful determinants such as education, employment, income, smoking and weight (see Figure 2). Remarkably, both studies gathered data on whether participants in their surveys were involved in choral singing, but both failed to find any evidence of a link between singing and the measures of health employed.

Both surveys were examined very closely and the conclusion was reached that both studies have methodological and reporting weaknesses and that the evidence reported cannot provide a credible test of the hypothesis underpinning this review. For the Bygren et al., study, the quality of information on the extent and nature of participants' involvement in choral singing is very weak, and the health outcome employed is extremely challenging i.e. whether a participant survives or dies in the course of the study. It is highly implausible that any relationships would be found, especially following statistical controls for a wide range of highly influential factors (see Table 24). The Hyyppä and Mäki study too was judged to have substantial weaknesses – not least that the proportions of choral singers in the sample were so low, especially for the men, that even simple order relationships between involvement in singing and the self-assessed measure of health, could not be demonstrated (see Table 25).

There is no question that in principle findings from large scale population surveys could be relevant to the hypothesis that group singing is beneficial for wellbeing and health. Nevertheless, the scale of the studies required, the care needed in gathering appropriately detailed and sensitive data, and most importantly, the current lack of a

convincing case for expecting substantial effects, render it very unlikely that specifically designed studies would ever be mounted.

#### **5.4 Theoretical perspectives on singing, wellbeing and health**

In the studies under review a number of theoretical perspectives and principles developed outside the specific field of singing and health are drawn upon to understand how singing could have benefits for wellbeing and health.

Bailey and Davidson (2002b, 2005) for example, make a convincing case for the relevance of Csikszentmihalyi's (1997) 'flow theory' in understanding the mechanisms through which singing can be beneficial to health. In so far as singing is a demanding and valued activity which requires concentration and practice, it is capable of creating a sense of 'absorption' or 'flow' allowing the singer to escape from other preoccupations. Bailey and Davidson also highlight the arguments put forward by Csikszentmihalyi for the importance of social interaction for mental wellbeing, and link this to the inherently social character of group singing.

Cohen et al. (2006) also draw upon general theoretical principles developed in the field of Gerontology, in attempting to explain why creative activities, including singing in groups, have potential wellbeing and health benefits for older people. The key processes being that creative activity helps to generate a 'sense of control' and provides opportunities for 'social engagement'. Interestingly, they also add a further factor of motivation which they see as generated by the intrinsic aesthetic qualities associated with creative activity.

#### **5.5 Conclusions and Recommendations**

In all human cultures and throughout history, music has played a significant role in human life. Music indeed can be regarded as one of the defining features of our human nature, and singing is the one form of musical expression that is freely open to everyone. Given the fact that group singing can be powerful and moving experience, and given the possibility that singing could contribute to quality of life, wellbeing and even health – it is remarkable that so little research has explored these issues, and disappointing that the research which has been undertaken is so limited and so variable in character.

A number of researchers/studies stand out as making important contributions to what is clearly a field of research in an early stage of development:

- Bailey and Davidson (2002b, 2005) have shown powerfully that amateur group singing can have benefits for participants across a wide social spectrum – from homeless men singing together, to middle class singers in traditional choral societies
- Cliff and Hancox (2001) have begun to explore in more detail the range of benefits experienced by choral singers, and have identified empirically a number of key dimensions of benefit which can be measured



- Beck et al. (2000), Grape et al. (2002) and Kreutz et al. (2004) have begun to explore the possible physiological mechanisms which could underpin subjective experiences in both individual and choral singing
- Silber (2005) has undertaken a detailed analysis of the specific contributions which musical structures and processes can make in processes of developing social and personal wellbeing in a challenging group of women prisoners
- Louhivouri et al.(2005) has linked research on choral singing to the wider issue of 'social capital' and begun to explore experiences and benefits of group singing in very contrasting cultural contexts
- Cohen et al. (2006, 2007) have conducted an ambitious quasi-experimental study of the possible effects of community singing on the mental and physical health of older people

At this stage, if work on the potential value of group singing for wellbeing and health is to develop, there is a need to create a collaborative and progressive programme embracing conceptual clarification / theory building and empirical research.

Conceptual clarification is needed with respect to the nature of singing itself, the various forms it can take, and the essential structures and processes involved (physically, physiologically and psychologically) in the production of the 'speaking voice' and the 'singing voice.'

Conceptual clarification is also needed with respect to the nature of wellbeing and health. The World Health Organisation has sponsored a large-scale international collaborative programme of work on health-related quality of life, which has elaborated and operationalised the WHO definition of health as 'a complete state of physical, mental and social wellbeing, and not merely an absence of illness or infirmity' (WHO, 1946). Conceptual and theoretical developments in the field of 'wellbeing' have seen a remarkable growth in recent years and the edited volume by Huppert, Baylis and Keverne (2005) represents a landmark text in this energetic area of work. Further work on singing, wellbeing and health needs to take note of these recent conceptual and theoretical developments.

Specific theoretical work is required to link the nature, process and experiences of 'singing' with the nature, processes and experiences of wellbeing and health. The key questions that need to be addressed are 'What?' and 'How?' What effects, if any, does active involvement in group singing have on wellbeing and health? And how does singing have these effects – what mechanisms are at work? What mechanisms, for example, link the physicality of singing with the physical wellbeing of the body? What mechanisms link the psychological and social processes at work during singing with our sense of emotional wellbeing and mental health? Crucial to the theoretical task is to identify clearly those mechanisms which are distinctively and uniquely at work in the process of singing – and those which are contingent and may be brought into play equally well by a range of other social or creative activities.

Conceptual and theoretical work should go hand in hand with a systematic programme of empirical research. Research can help to inform theory development

up to a point, but a truly progressive research programme requires the elaboration of a realistic theoretical framework which can generate hypotheses. A progressive programme also requires a critical mass of researchers and research teams working in collaboration and competition in testing hypotheses and establishing robust research findings through well-designed studies, replication and synthesis of evidence. As yet the field is a long way from establishing even the foundation for such a progressive programme, but without it, future research is destined to be haphazard, unconnected and non-cumulative.

It is hoped that this review will contribute to a process of developing the progressive research programme needed to take this area of work forward.

## 6. REFERENCES

### 54 reports identified for the review

- Adams, K. (2000) How can I keep from singing: a lifetime of sweet singing in the choir, in B.A. Roberts and A. Rose (eds.) *Sharing the Voices: The Phenomenon of Singing 2*. Proceedings of the International Symposium, St. John's, Newfoundland, Canada, 1999. St. John's: Memorial University of Newfoundland.
- Anshel, A. and Kipper, D.A. (1988) The influence of group singing on trust and cooperation, *Journal of Music Therapy*, XXV(3), pp.145-155.
- Ashley, M. (2002) Singing, gender and health: perspectives from boys singing in a church choir, *Health Education*, 102(4), pp.180-186.
- Bailey, B.A. and Davidson, J.W. (2002a) Emotional, social, and cognitive enrichment through participation in group singing: interviews with members of a choir for homeless men, in A. Rose, K. Adams and R. Chisholm (eds.) *Sharing the Voices: The Phenomenon of Singing III*. Proceedings of the International Symposium, St. John's, Newfoundland, Canada, 2001. St. John's: Memorial University of Newfoundland.
- Bailey, B.A. and Davidson, J.W. (2002b) Adaptive characteristics of group singing: perceptions from members of a choir for homeless men, *Musicae Scientiae*, VI(2), pp.221-256.
- Bailey, B.A. and Davidson, J.W. (2003a) Perceived holistic health effects of three levels of music participation, in R. Kopiez, A.C. Lehmann, I. Wohther, and C. Wolf (eds.) *Proceedings of the 5<sup>th</sup> Triennial ESCOM Conference*, 8-13 September 2003, Hanover University of Music and Drama, Germany.
- Bailey, B.A. and Davidson, J.W. (2003b) Amateur group singing as a therapeutic instrument, *Nordic Journal of Music Therapy*, 12(1), pp.18-33.
- Bailey, B.A. and Davidson, J.W. (2005) Effects of group singing and performance for marginalized and middle-class singers, *Psychology of Music*, 33(3), pp.269-303.
- Beck, R.J., Cesario, T.C., Yousefi, A. and Enamoto, H. (2000) Choral singing, performance perception, and immune system changes in salivary immunoglobulin A and cortisol, *Music Perception*, 18(1), pp.87-106.
- Beck, R.J., Gottfried, T.L., Hall, D.J., Cisler, C.A. and Bozeman, K.W. (2006) Supporting the health of college solo singers: the relationship of positive emotions and stress to changes in salivary IgA and cortisol during singing, *Journal of Learning through the Arts: A Research Journal on Arts Integration in Schools and Communities*, 2(1), article 19.
- Bygren, L.O., Konlaan, B.K. and Johansson, S-E. (1996) Attendance at cultural events, reading books or periodicals, and making music or singing in a choir as

determinants for survival: Swedish interview survey of living conditions, *British Medical Journal*, 313, pp.1577-1580.

Clift, S.M. and Hancox, G. (2001) The perceived benefits of singing: findings from preliminary surveys of a university college choral society, *Journal of the Royal Society for the Promotion of Health*, 121(4), pp.248-256.

Chorus America (2003) *America's Performing Art: A study of choruses, choral singers, and their impact*, Washington: Chorus America.

Available from: <http://www.chorusamerica.org/files/chorstudy.pdf>

Cohen, G.D., Perlstein, S., Chapline, J., Kelly, J., Firth, K.M. and Simmens, S. (2006) The impact of professionally conducted cultural programs on the physical health, mental health, and social functioning of older adults, *The Gerontologist*, 46(6), pp.726-734.

Cohen, G.D., Perlstein, S., Chapline, J., Kelly, J., Firth, K.M. and Simmens, S. (2007) The impact of professionally conducted cultural programs on the physical health, mental health and social functioning of older adults – 2-year results, *Journal of Aging, Humanities and the Arts*, 1, pp.5-22.

Cusicanqui, M. and Salmon, R. (2004) Seniors, small fry, and song: a group work libretto of an intergenerational singing group, *Journal of Gerontological Social Work*, 44(1-2), pp.189-210.

Durrant, C. and Himonides, E. (1998) What makes people sing together? Socio-psychological and cross-cultural perspectives on the choral phenomenon, *International Journal of Music Education*, 32, pp.61-70.

Faulkner, R. and Davidson, J.W. (2004) Men's vocal behaviour and the construction of self, *Musicae Scientiae*, 8(2), pp.231-255.

Faulkner, R. and Davison, J.W. (2006) Men in chorus: collaboration and competition in homo-social vocal behaviour, *Psychology of Music*, 34(2), pp.219-237.

Galati, D., Costa, T., Rognoni, E., Pisterzi, A. (2006) The effect of singing on subjective emotional state and psychophysiological response in singers, paper presented at the 9<sup>th</sup> *International Conference on Music Perception and Cognition*, Alma Mater Studiorum, University of Bologna, 22-26 August, 2006.

Gould, W.J. and Okamura, H. (1973) Static lung volumes in singers, *Annals of Otology, Rhinology and Laryngology*, 82, pp.89-95.

Grape, C., Sandgren, M., Hansson, L-O., Ericson, M. and Theorell, T. (2003) Does singing promote well-being? An empirical study of professional and amateur singers during a singing lesson, *Integrative Physiological and Behavioral Science*, 38(1), pp.65-74.

Heller, S.S., Hicks, W.R. and Root, W.S. (1960) Lung volumes of singers, *Journal of Applied Physiology*, 15(1), pp.40-42.

- Hillman, S. (2002) Participatory singing for older people: a perception of benefit, *Health Education*, 102(4), pp.163-171.
- Hills, P. and Argyle, M. (1998a) Musical and religious experiences and their relationship to happiness, *Personality and Individual Differences*, 25, pp.91-102.
- Hills, P. and Argyle, M. (1998b) Positive moods derived from leisure and their relationship to happiness and personality, *Personality and Individual Differences*, 25, pp.523-535.
- Houston, D.M., McKee, K.J., Carroll, L. and Marsh, H. (1998) Using humour to promote psychological wellbeing in residential homes for older people, *Aging and Mental Health*, 2(4), pp.328-332.
- Hyppä, M.T and Mäki, J. (2001) Individual-level relationships between social capital and self-rated health in a bilingual community, *Preventive Medicine*, 32, pp.148-155.
- Johansson, S.E., Konlaan, B.B. and Bygren, L.O. (2001) Sustaining habits of attending cultural events and maintenance of health: a longitudinal study, *Health Promotion International*, 16(3), pp.229-234.
- Joyce, M. (1996) Turn off the radio and sing for your lives! Women, singing, and experiential education, In K. Warren (ed.) *Women and Experiential Education*, Dubuque: Kendall/Hunt.
- Konlaan, B.B., Bygren, L.O. and Johansson, S-E. (2000) Visiting the cinema, concerts, museums or art exhibitions as determinant of survival: a Swedish fourteen-year cohort follow-up, *Scandinavian Journal of Public Health*, 28(3), pp.174-8.
- Kreutz, G., Bongard, S., Rohrman, S., Grebe, D., Bastian, H.G. and Hodapp, V. (2003) Does singing provide health benefits? in R. Kopiez, A.C. Lehmann, I. Wohther, and C. Wolf (eds.) *Proceedings of the 5<sup>th</sup> Triennial ESCOM Conference*, 8-13 September 2003, Hanover University of Music and Drama, Germany.
- Kreutz, G., Bongard, S., Rohrman, S., Grebe, D., Bastian, H.G. and Hodapp, V. (2004) Effects of choir singing or listening on secretory immunoglobulin A, cortisol and emotional state, *Journal of Behavioral Medicine*, 27(6), pp.623-635.
- Kuhn, D. (2002) The effects of active and passive participation in musical activity on the immune system as measured by salivary immunoglobulin A (SigA), *Journal of Music Therapy*, 39(1), pp.30-39.
- Kennedy, M.A. (2002) 'It's cool because we like to sing:' junior high school boys' experience of choral music as an elective, *Research Studies in Music Education*, 18, pp.26-37.
- Louhivuori, J. (2006) Cross-cultural approach to emotions in choir singing, paper presented at the 9<sup>th</sup> *International Conference on Music Perception and Cognition*, Alma Mater Studiorum, University of Bolgna, 22-26 August, 2006.

Louhivuori, J., Salminen, V-M. and Lebaka, E. (2005) "Singing Together" – A cross-cultural approach to the meaning of choirs as a community, In Campbell, P.S., Drummond, J., Dunbar-Hall, P., Howard, K., Schippers, H. and Wiggins, T. (eds.) (2005) *Cultural Diversity in Music Education: Directions and Challenges for the 21<sup>st</sup> Century*, Australian Academic Press.

Michalos, A.C. (2005) Arts and the quality of life: an exploratory study, *Social Indicators Research*, 71, pp.11-59.

Montemurro, R.N.R. (1996) Singing lullabies to unborn children: experiences in Village Vilamarxant, Spain, *Pre- and Perinatal Psychology Journal*, 11(1), pp.9-16.

Patteson, A. (2000) Singing a woman's life: how singing lessons transformed the lives of nine women, in B.A. Roberts and A. Rose (eds.) *Sharing the Voices: The Phenomenon of Singing 2*. Proceedings of the International Symposium, St. John's, Newfoundland, Canada, 1999. St. John's: Memorial University of Newfoundland.

Pitts, S.E. (2004) 'Everybody wants to be Pavarotti': the experience of music for performers and audience at a Gilbert and Sullivan festival, *Journal of the Royal Musical Association*, 129, pp.149-167.

Richards, H. and Durrant, C. (2003) To sing or not to sing: a study on the development of 'non-singers' in choral activity, *Research Studies in Music Education*, 20, pp.78-89.

Rider, M., Mickey, C., Weldin, C. and Hawkinson, R. (1991) The effects of toning, listening and singing on psychophysiological responses, in C.D. Maranto (ed.) *Applications of Music in Medicine*, Washington: National Association of Music Therapy.

Sandgren, M. (2002) Voice, soma, and psyche: a qualitative and quantitative study of opera singers, *Medical Problems of Performing Artists*, 17, pp.11-21.

Schorr-Lesnack, B., Teirstein, A.S., Brown, L.K. and Miller, A. (1985) Pulmonary function in singers and wind-instrument players, *Chest*, 88(2), pp.201-205.

Silber, L. (2005) Bars behind bars: the impact of a women's prison choir on social harmony, *Music Education Research*, 7(2), pp.251-271.

Silvey, P.E. (2002) Perspectives of aging adult choral musicians: implications for meaningful lifelong participation in ensemble singing, in A. Rose, K. Adams and R. Chisholm (eds.) *Sharing the Voices: The Phenomenon of Singing III*. Proceedings of the International Symposium, St. John's, Newfoundland, Canada, 2001. St. John's: Memorial University of Newfoundland.

Smithrim, K.L. (1998) Still singing for our lives: singing in the everyday lives of women through this century, in B.A. Roberts (ed.) *Sharing the Voices: The Phenomenon of Singing*. Proceedings of the International Symposium, St. John's, Newfoundland, Canada, 1997. St. John's: Memorial University of Newfoundland.

Steurer, M., Simak, S., Denk, D.M. and Kautzky, M. (1998) Does choir singing cause noise-induced hearing loss? *Audiology*, 37(1), pp.38-51.

Unwin, M.M., Kenny, D.T. and Davis, P.J. (2002) The effects of group singing on mood, *Psychology of Music*, 30, pp.175-185.

Valentine, E. and Evans, C. (2001) The effects of solo singing, choral singing and swimming on mood and physiological indices, *British Journal of Medical Psychology*, 74, pp.115-120.

Watanabe, H. (2005) Changing adult learning in Japan: the shift from traditional singing to karaoke, *International Journal of Lifelong Education*, 24(3), pp.257-267.

Windsor, J. (2005) *Your Health and the Arts: A Study of the Association between Arts Engagement and Health*, London: Arts Council England.

Wise, G.W., Hartmann, D.J. and Fisher, B.J. (1992) Exploration of the relationship between choral singing and successful aging, *Psychological Reports*, 70, pp. 1175-1183.

### **Additional references**

Batt-Rawden, K.B., DeNora, T. and Ruud, E. (2005). Music listening and empowerment in health promotion; A study of the role and significance of music in everyday life of the long-term ill. *Nordic Journal of Music Therapy*, 14(2), pp.120-136.

Beck, T.A., Mendes de Leon, C., Marottoli, A. and Berkman, L. (1999) Population based study of social and productive activities as predictors of survival among elderly Americans, *British Medical Journal*, 319, pp.478-483.

Brown, S. And Volgsten, U (2006) *Music and Manipulation: On the social uses and social control of music*, Oxford: Blackwell.

Cassidy J W. and Standley J M. (1995) The effect of music listening on physiological responses of premature infants in the NICU, *Journal of Music Therapy*, 32(4),pp. 208-227.

Clift, S. (2002) *Evaluation of the SoundStart Project, Isle of Wight*. Canterbury: Canterbury Christ Church University College, unpublished report.

Clift, S., Hancox, G., Morrison, I., Hess, B., Kreutz, G. and Stewart, D. (2007) Choral Singing and Psychological Wellbeing: Findings from English choirs in a cross-national survey using the WHOQOL-BREF, In Williamon, A. and Coimbra, D. (Eds.) *Proceedings of the International Symposium on Performance Science*, Porto, November 2007.

Clift, S. and Hancox, G. (2008) The Significance of Choral Singing for Sustaining Psychological Wellbeing: Findings from a survey of choristers in England, Germany

and Australia, Paper presented at the *International Society for Music Education, Community Music and Wellbeing Commission Seminar*, Rome, July 2008.

Clift, S., Mackenzie, K. and Bushell, F. (2006) *The MusicStart Project: Evaluation of an Early Years Music Initiative*. Sidney De Haan Report 1, Canterbury: Canterbury Christ Church University.

Clift, S. and Vella-Burrows, T. (2004) *Arts and Healthy Communities in the South East of England*. Canterbury: Canterbury Christ Church University College.

Coffman, D. (2002) Music and quality of life in older adults, *Psychomusicology*, 18, pp.76-88.

Csikszentmihalyi, M. (2002) *Flow*, London: Rider

DeNora, T. (2000) *Music in Everyday Life*. Cambridge: Cambridge University Press.

Duchen, J. (2008) Mind the Bach, *The Independent*, Wednesday 26 March 2008 (Extra, p.14)

Everitt, A. (1997) *Joining In: An investigation into participatory music*. London: Calouste Gulbenkian Foundation.

Gouk, P. (Ed.) (2000) *Musical Healing in Cultural Contexts*, Aldershot: Ashgate.

Horden, P. (Ed) (2000) *Music as Medicine: The history of music therapy since antiquity*. Aldershot. Ashgate.

Hays, T. and Minichiello, V. (2005) The contribution of music to quality of life in older people: An Australian qualitative study, *Ageing & Society*, 25, pp.261-278.

Hyypä, M.T., Mäki, J., Impivaara, O. and Aromaa, A. (2005) Leisure participation predicts survival: a population-based study in Finland, *Health Promotion International*, 21(1), pp.5-12.

Ionnides, J.P.A. (2005) Why most published research findings are false, *PLoS Medicine*, 2(8), pp.696-701. (Available from: [www.plosmedicine.org](http://www.plosmedicine.org) )

Juslin, P.K. and Västfjäll, D. (2008) Emotional responses to music: the need to consider underlying mechanisms, *Behavioural and Brain Sciences* (in press).

Juslin, P.K. and Sloboda, J. (eds.)(2001) *Music and Emotion: Theory and research*, Oxford: Oxford University Press.

Kashdan, T.B. (2004) The assessment of subjective well-being (issues raised by the Oxford Happiness Questionnaire), *Personality and Individual Differences*, 36, pp.1225-1232.



Konlaan, B.B., Björby, N., Bygren, L.O., Weissglas, G., Karlsson, L.G. and Widmark, M. (2000) Attendance at cultural events and physical exercise and health: a randomised controlled study, *Public Health*, 114, pp.316-319.

Konlaan, B.B., Theobald, H. and Bygren, L-O. (2002) Leisure time activity as a determinant of survival: a 26-year follow-up of a Swedish cohort, *Public Health*, 116, pp.227-230.

Kototsaki, D. and Hallam, S. (2007) Higher education music students' perceptions of the benefits of participative music making, *Music Education Research*, 9(1), pp.93-109.

Lecanuet, J-P. (1996) Prenatal auditory experience. In I. Deliège and J. Sloboda (Eds.) *Musical Beginnings: origins and development of musical competence*. Oxford: Oxford University Press. (pp. 3-34)

Lesta, B. and Petocz, P. (2006) Familiar group singing: Addressing mood and social behaviour of residents with dementia displaying sundowning, *Australian Journal of Music Therapy*, 17. (Available from: <http://www.austmta.org.au/> )

Levitin, D. (2006) *This is Your Brain on Music*. London: Atlantic Books.

Mithen, S. (2005) *The Singing Neanderthals*. London: Weidenfeld and Nicholson.

Pavlicevic, M. and Ansdell, G (Eds.) (2004) *Community Music Therapy*. London: Jessica Kingsley Publishers.

Papousek M (1996) Intuitive parenting: a hidden source of musical stimulation in infancy. In I. Deliège and J. Sloboda (Eds.) *Musical Beginnings: origins and development of musical competence*. Oxford: Oxford University Press. (pp.88-114)

Petticrew, M. and Roberts, H. (2003) Evidence, hierarchies, and typologies: horses for courses, *Journal of Epidemiology and Community Health*, 57, 527-529.

Petticrew, M. and Roberts, H. (2006) *Systematic Reviews in the Social Sciences: A practical guide*. Oxford: Blackwell.

Robinson, S. and Clift, S. (2002) *Arts, Wellbeing and Health*, Canterbury: Canterbury Christ Church University College.

Rohwer, D. and Coffman, D. (2006) Relationships between wind band membership, activity level, spirituality, and quality of life in older adults, *Research Perspectives in Music Education*, 10, pp.21-27.

Särkämö, T., Tervaniemi, M., Laitinen, S., Forsblom, A., Soinila, S., Mikkonen, M., Autti, T., Silvennoinen, H.M., Erkkilä, J., Laine, M., Peretz, I. and Hietanen, M. (2008) Music listening enhances cognitive recovery and mood after middle cerebral artery stroke, *Brain*, 131, pp.866-876.

Sterne, J.A.C. and Davey Smith, G. (2001) Sifting the evidence – what's wrong with significance tests? *British Medical Journal*, 322, pp.226-231.

Titze, I.R. (2008) The human instrument, *Scientific American*, 298(1), pp.78-85.

Trehub S E and Schellenberg E G (1995) Music: its relevance to infants. In R. Vasta (Ed.) *Annals of Child Development*, 11, pp.1-24.

Trehub S E and Trainor L (1998) Singing to infants: lullabies and play songs. *Advances in Infancy Research*, 12, pp.43-77.

Trevarthen, C. (1999-2000) Musicality and the intrinsic motive pulse: evidence from human psychobiology and infant communication. *Musicae Scientiae*, Special Issue (1999-2000), pp.155-215

WHO (1946) Preamble to the Constitution of the World Health Organization as adopted by the International Health Conference, New York, 19-22 June, 1946; signed on 22 July 1946 by the representatives of 61 States (*Official Records of the World Health Organization*, no. 2, p.100) and entered into force on 7 April 1948.

Wilkinson, A.V., Waters, A.J., Bygren, L.O. and Tarlov, A.R. (2007) Are variations in rates of attending cultural activities associated with population health in the United States? *BMC Public Health*, 7, pp.226-233.

## APPENDIX 1

### The systematic search process, February 2007

#### Christine Whitmore

A comprehensive search of the literature requires the effective use of the most appropriate database(s). In order to retrieve research articles about singing and health, the following literature databases were searched:

- Medline
- Embase
- Cinahl
- PsychInfo
- ERIC
- Cochrane Library
- ASSIA
- RILM
- Music Therapy World

Many databases, such as Medline, use a thesaurus (a controlled vocabulary) for indexing articles to enable more effective searching. The vocabulary is controlled by grouping related concepts under a single preferred term. A thesaurus will typically contain indexed preferred terms (or 'descriptors'), definitions of those preferred terms, and cross-references between preferred terms. Medline and the Cochrane Library both use a thesaurus called MeSH (Medical Subject Headings).

In our literature searches, search terms for each database were selected mostly along the lines of terms used to index six articles we considered to be key to the review at the outset (Bailey and Davidson, 2005; Beck et al., 2000; Bygren, Konlaan and Johansson, 1996; Clift and Hancox, 2001, Grape et al., 2003 and Kreutz, et al., 2004) in each of those databases. None of the 'example articles' were indexed in Cinahl, ERIC or ASSIA databases so these sources were searched using terms similar to those used in Medline, Embase and PsychInfo. Selected search terms conformed to the controlled vocabulary for each database.

Free text terms were used when the controlled vocabulary was not sufficiently expressive. The concept of 'singing' was seen as an essential element in the studies to be included in this systematic review. Since 'singing' is not a preferred term in the MeSH controlled vocabulary, the free text term 'singing' was used with controlled vocabulary terms in searching databases using MeSH (Medline, Cochrane).<sup>1</sup> 'Music therapy' was not an index term for any of our 'example articles' but we found it absolutely necessary to include it in our search strategies because music therapy often described the tone and flavour of the type of research we were interested in.

---

<sup>1</sup> Although keyword searching using MeSH terms is a more precise way of searching, this is not always possible. There may be times when there is no MeSH preferred term for the subject being searched; for example, until 1997, there was no MeSH preferred term for "evidence based medicine". In the absence of MeSH preferred terms, one needs to use free text terms for searching (SchARR, U of Sheffield, <http://www.shef.ac.uk/scharr/reswce/litsear.htm> accessed 1.8.06)

We searched music therapy and singing separately and then combined the results with singing and health-related terms.

Search strategies were iteratively developed with discussions about which searches provided the right amount of sensitivity<sup>2</sup> and specificity.<sup>3</sup> In systematic reviews specificity is often sacrificed in favour of sensitivity to maximize the yield of relevant articles.<sup>4</sup> This means that large numbers of bibliographic references were retrieved for consideration for inclusion in the review.

The following are details of the searches done on the chosen databases:

---

<sup>2</sup> Sensitivity (recall) is the proportion of all relevant studies in the database that a search retrieves

<sup>3</sup> Specificity (precision) is the proportion of all studies retrieved in a search that are relevant

<sup>4</sup> Centre for Reviews and Dissemination - Undertaking Systematic Reviews of Research on Effectiveness- CRD's Guidance for those Carrying Out or Commissioning Reviews CRD Report Number 4 -2nd Edition- March 2001

## Medline

On 19 February 2007 the Medline database (1950 to date) was searched using the Web-based Datastar platform, with no date or language constraints.

Number	Search Term	Results
1	singing	1126
2	Music.W..DE.	6655
3	Voice.W..DE.	5100
4	vocal	16512
5	3 OR 4	20076
6	2 AND 5	374
7	1 OR 6	1351
8	Music-Therapy.DE.	1371
9	Health#.W..DE	171189
10	Quality-Of-Life.DE	58634
11	Personal-Satisfaction.DE	5491
12	Life-Style#.DE	38188
13	well ADJ being OR wellbeing OR well-being	23964
14	Leisure-Activities#.DE.	87382
15	Affect.W..DE.	14446
16	9 OR 10 OR 11 OR 12 OR 13 OR 14 OR 15	372829
17	7 AND 16	54
18	7 AND 8	38
19	(17 OR 18) AND HUMAN=YES	82

Singing was searched as a free text term because it was not part of the controlled vocabulary. Since singing was not an index term in this database, the terminology representing singing was broadened to include the index term music ANDed (intersected) with the index term, voice, OR the free text term, vocal. Well-being (and its various spellings) was searched as a free text term because it was not a part of the controlled vocabulary. Music therapy, quality of life and personal satisfaction were the lowest terms on their hierarchical trees and could not be exploded. Affect was not exploded because 'irritable mood' was the only narrower term below and it was not relevant. Health, life-style and leisure activities were exploded because this would retrieve articles indexed to relevant narrower terms in their trees. Singing was 'ANDed' (intersected) with music therapy separately so that we could keep a close eye on what type of articles this term was retrieving as opposed to singing 'ANDed' (intersected) with the health, quality of life, personal satisfaction, life style, well-being, leisure activities and affect terms. We were satisfied with the two sets (#17, #18), so they were combined. Results were limited to human-related literature.

## Embase

On 19 February 2007 the Embase database (1974 to date) was searched using the Web-based Datastar platform, with no date or language constraints.

Number	Search Term	Results
1	Singing.W..DE	1289
2	Music-Therapy.DE.	1200
3	Health#.W..DE.	87081
4	Quality-Of_Life#.DE.	76214
5	Wellbeing.W..DE.	13713
6	Satisfaction.W..DE. OR Life-Satisfaction.DE.	8311
7	Social-Aspect.DE.	46468
8	Emotion#.W..DE OR Emotionality.W..DE.	127295
9	Affect.W..DE.	5763
10	Lifestyle.W..DE.	26014
11	3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9 OR 10	339838
12	1 AND 2	19
13	1 AND 11	74
14	(12 OR 13) AND HUMAN=YES	74

Singing, music therapy, well-being, social aspect, lifestyle and emotionality were all lowest terms on their hierarchical trees so they could not be exploded. Health was exploded because it is a broad term with relevant narrower terms such as health status and mental health below it. Quality of life was exploded because the terms beneath included quality of life indexes which might include singing. Satisfaction and life satisfaction were included but other satisfactions such as job satisfaction and sexual satisfaction were not included. Emotion was exploded because it included all the variations on emotion such as mood and happiness. Affect was not exploded because it did not include relevant narrow terms. Singing was 'ANDed' (intersected) with music therapy separately so that we could see the type of articles retrieved, and then combined with the results of singing 'ANDed' (intersected) with health, quality of life, well-being, satisfaction, life satisfaction, social aspect, emotion, emotionality, affect and lifestyle (#12, #13). Results were limited to human-related literature.

## Cinahl

On 19 February 2007 the Cinahl database (1982 to date) was searched using the Web-based Datastar platform, with no date or language constraints.

Number	Search Term	Results
1	Singing.W..DE	266
2	Music-Therapy.DE.	1044
3	Health#.W..DE.	67914
4	Quality-Of_Life#.DE.	17031
5	Personal-Satisfaction.DE.	1580
6	Life-Style#.DE.	40867
7	Psychological-Well-Being.DE	2771
8	well ADJ being OR wellbeing OR well-being	18584
9	Group-Processes.DE	2090
10	Emotions#.W..DE	19955
11	3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9 OR 10	145953
12	1 AND 2	15
13	1 AND 11	29
14	12 OR 13	39

Singing, music therapy, personal satisfaction and psychological well-being were all lowest terms on their hierarchical trees, thus could not be exploded. The various spellings of well-being were searched as free-text terms. The term 'group processes' was kept at its broadest level (as it was in one of our 'example' articles) while health, quality of life, life-style and emotions were exploded because there were relevant narrower terms below them. Singing was 'ANDed' (intersected) with music therapy separately to control the type of articles retrieved, and then combined with the results of singing 'ANDed' (intersected) with health, quality of life, personal satisfaction, life-style, psychological well-being, the various spellings of well-being, group processes and emotions (#12, #13). No limiting to humans was available.

## PsychInfo

On 19 February 2007 the PsychInfo database (1806 to date) was searched using the Web-based Datastar platform, with no date or language constraints.

Number	Search Term	Results
1	Singing.W..DE	295
2	Music-Therapy.DE.	1730
3	Health#.W..DE	40690
4	Quality-Of-Life#.DE	11829
5	Psychosocial-Factors.DE	17959
6	Satisfaction.W..DE. OR Life-Satisfaction.DE.	9415
7	Well-BeingDE.	9501
8	Social-Interaction.DE. OR Group-Performance.DE	13939
9	Emotional-States.DE	18060
10	Emotional-Responses.DE	9945
11	3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9 OR 10	119148
12	1 AND 2	23
13	1 AND 11	22
14	12 OR 13	38

Singing, music therapy, psychosocial factors, satisfaction, life satisfaction, well-being and group performance were all lowest terms on their hierarchical trees so they could not be exploded. The terms social interaction, emotional responses and emotional states were kept at their broadest level because exploding them would have brought in too many irrelevant articles. Health and quality of life were exploded because they had relevant narrower terms below them. Singing was 'ANDed' (intersected) with music therapy separately to see the type of articles retrieved and then combined with the results of singing 'ANDed' (intersected) with health, quality of life, psychosocial factors, satisfaction, life satisfaction, well-being, social interaction, group performance, emotional states, emotional responses (#12, #13). No limiting to humans was available.



## ERIC

On 19 February 2007 the ERIC database (1966 to date) was searched using the Web-based Datastar platform, with no date or language constraints.

Number	Search Term	Results
1	Singing.W..DE.	703
2	Music-Therapy.DE.	153
3	Health#.W..DE. OR Health-Activities#.DE.	20215
4	Quality-Of-Life#.DE.	12151
5	Satisfaction.W..DE. OR Community-Satisfaction.DE. OR Life-Satisfaction.DE. OR Participant-Satisfaction.DE.	5671
6	Affective-Behavior.DE. OR Emotional-Experience.DE. OR Emotional-Response.DE.	8153
7	Group-Activities.DE. OR Group-Experience.DE.	3786
8	3 OR 4 OR 5 OR 6 OR 7	46642
9	1 AND 2	4
10	1 AND 8	35
11	9 OR 10	38

Singing and music therapy were both lowest terms on their hierarchical trees so they could not be exploded. Satisfaction was partially broadened to include community satisfaction, life satisfaction and participant satisfaction. Health was exploded as was the term, health activities, which included the narrower term, health promotion. Affective behavior, emotional response, group activities and group experience were all narrowest terms and emotional experience did not have any relevant narrower terms. Quality of life was exploded to include the narrower term, well-being. Singing was 'ANDed' (intersected) with music therapy separately to keep an eye on the type of articles retrieved, and then this set was combined with the results of singing 'ANDed' with health, health activities, quality of life, the various satisfactions, affective behavior, emotional experience, emotional response, group activities, group experience (#9, #10). No limiting to humans was available.

## The Cochrane Library

The full collection of databases comprising the Cochrane Library on the Web-based Wiley InterScience platform - with no date or language constraints - was searched on 19 February 2007.

Number	Search Term	Results
1	singing	302
2	Music.W..DE.	192
3	Voice.W..DE.	52
4	vocal	474
5	3 OR 4	515
6	2 AND 5	6
7	1 OR 6	306
8	Music-Therapy.DE.	208
9	Health#.W..DE	2983
10	Quality-Of_Life.DE	6929
11	Personal-Satisfaction.DE	155
12	Life-Style#.DE	1062
13	well ADJ being OR wellbeing OR well-being	57051
14	Leisure-Activities#.DE.	4431
15	Affect.W..DE.	2003
16	9 OR 10 OR 11 OR 12 OR 13 OR 14 OR 15	70469
17	7 AND 16	84
18	7 AND 8	12
19	17 OR 18	85

Singing and well-being (and its various spellings) were searched as free text terms because they were not a part of the controlled vocabulary (MeSH). Since singing was not an index term in Cochrane, the terminology representing singing was broadened to include the index term music ANDed (intersected) with the index term voice OR the free text term vocal. Music therapy, quality of life and personal satisfaction were the lowest terms on their hierarchical trees so they could not be exploded. Affect was not exploded because the narrower term was not relevant. Health, life-style and leisure activities were exploded because this would retrieve articles indexed to relevant narrower terms in their trees. Singing was 'ANDed' (intersected) with music therapy separately so that we could keep a close eye on what type of articles this term was retrieving as opposed to singing 'ANDed' with the health, quality of life, personal satisfaction, life style and well-being terms. We were satisfied with the two sets (#17, #18), so they were combined. No limiting to humans was available.

### **Assia**

The ASSIA (Applied Social Science Index and Abstracts) database (1987 to date) on the CSA Internet Database Service platform - with no date or language constraints - was searched on 19 February 2007.

Number	Search Term	Results
1	Singing.DE	13
2	singing	69
3	1 OR 2	69

The preferred indexing term 'singing' was 'ORed' (combined) with the free text term 'singing' in order to increase results. The search in ASSIA was not filtered further because of the small number of results using this most general term. Limiting to humans was not available.

### **RILM**

The RILM Abstracts of Music Literature database (1967 to date) on the Web-based OCLC First Search platform - with no date or language constraints - was searched on 14 October 2006.

Number	Search Term	Results
1	(kw: singing AND kw: health) OR (kw: singing AND kw: music w therapy)	196

RILM uses a keyword index which includes words from titles, subject headings, and notes. In this database, it was decided not to elaborate on our basic 'health' search term because broadening in that area brought no new results of interest. Singing was 'ANDed' (intersected) with health and combined with the results from intersecting singing with music therapy – resulting in 196 citations from the literature.

### **Music Therapy World**

The Music Therapy World Web site, based at the University of Witten/Herdecke in Germany, provides access to a database of references (entitled the 'Bibliography' database) from the music therapy literature compiled from a variety of sources including music therapy journal indices. This database was searched on 14 October 2006 using the term 'singing' only because the database does not have functionality for multiple search terms and combinations of terms. This search resulted in ca. 400 citations from the literature.

### **Google Scholar**

Results of searching singing and health in this search engine resulted in numerous hits, most of which were irrelevant or already known to us.

### **Further searching**

After locating potentially relevant papers through database searching, the following actions took place:

- hard or electronic copies of relevant articles were obtained and references were checked
- citation indexes were consulted to identify any papers that cite our foundation articles as well as any key papers identified through the database searches
- authors working in the health and singing topic area were identified and contacted by email or post in order to identify any further studies or publications
- key journals were manually checked at the level of the contents page
- music and health research websites were monitored for evidence of research not yet identified.

All of these searches were systematic and documented.

## APPENDIX 2

### Categories for coding of papers identified in searches based on titles and abstracts

1. Empirical study of any design (quantitative or qualitative), which directly explores the effects of:
  - active participation in singing (solo or group),
  - as a sole activity, or as part of a wider range of activities,
  - on aspects of health, well-being, quality of life or positive biochemical markers,
  - conducted on non-clinical samples.

Highest relevance to the systematic review

2. Empirical evaluation of any design (quantitative or qualitative) which explores the effects of
  - a therapeutic intervention involving active participation in singing (solo or group), as a sole activity,
  - on aspects of a specific condition or more general aspects of health, well-being, quality of life or positive biochemical markers,
  - conducted with clinical samples by a music therapist.

Very relevant to the systematic review

3. Empirical evaluation of any design (quantitative or qualitative) which explores the effects of
  - a therapeutic intervention involving active participation in singing (solo or group), as part of a wider range of activities,
  - on aspects of a specific condition or more general aspects of health, well-being, quality of life or positive biochemical markers,
  - conducted with clinical samples by a music therapist.

Possible relevance to the systematic review

4. Empirical studies concerned with aspects of singing and health/well-being, health/social care or therapeutic interventions, other than those covered by 1 and 2 above (e.g. effects of being sung to in the context of care or as part of therapy, attitudes towards the value of singing, songs as a medium for health education, studies of the vocal health of singers).

May have some relevance

5. Empirical studies or evaluations concerned with singing as an educational/therapeutic intervention with children in educational settings (e.g. focus on learning difficulties, emotional & behavioural difficulties, relationship problems, communication difficulties)

Little or no relevance to the systematic review

6. Reviews of empirical studies on the effects of singing – especially those in categories 1 and 2 above.

Very relevant to the systematic review

7. Theoretical/discussion papers on singing or music more generally and its benefits (i.e. not a specific empirical study and not a review of studies).

Possible relevance in a contextual sense to the systematic review

8. All other papers that do not fall into one of the seven categories above are unlikely to have any clear relevance to the issues of interest. Most of these may well be clearly irrelevant ('singing' may be used in a figurative way in the title or abstract e.g. 'Are we singing from the same song sheet in elderly care?')

Irrelevant to the systematic review

## ***The research team***

Stephen Clift  
Grenville Hancox  
Rosalia Staricoff  
Christine Whitmore  
Ian Morrison  
Matthew Raisbeck

Further information about the work of the Sidney De Haan Research Centre for Arts and Health, see:

***[www.canterbury.ac.uk/centres/sidney-de-haan-research/index.asp](http://www.canterbury.ac.uk/centres/sidney-de-haan-research/index.asp)***

