Sex differences in frailty: A systematic review and meta-analysis

EH Gordon a*, NM Peel b, M Samanta c, O Theou d, SE Howlett d, RE Hubbard b

a School of Medicine, The University of Queensland, Brisbane, QLD, Australia
b Centre for Research in Geriatric Medicine, The University of Queensland, Brisbane, QLD, Australia
c QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia
d Department of Medicine, Dalhousie University, Halifax, Nova Scotia, Canada

* Corresponding author
Postal Address: Centre for Research in Geriatric Medicine, The University of Queensland, Level 2 Building 33 Princess Alexandra Hospital, Wolloongabba, QLD 4102, Australia
E-mail address: e.gordon@uq.edu.au
Abstract

(249 words)

Background

It is a well-described clinical phenomenon that females live longer than males, yet tend to experience greater levels of co-morbidity and disability. Females can therefore be considered both more frail (because they have poorer health status) and less frail (because they have a lower risk of mortality). This systematic review aimed to determine whether this ageing paradox is demonstrated when the Frailty Index (FI) is used to measure frailty.

Methods

Medline, EMBASE and CINAHL databases were searched for observational studies that measured FI and mortality in community-dwellers over 65 years of age. In five-year age groups, meta-analysis determined the sex differences in mean FI (MD = mean FI_{female} – mean FI_{male}) and mortality rate.

Results

Of 6482 articles screened, seven articles were included. Meta-analysis of data from five studies (37 426 participants) found that MD values were positive (p <0.001; MD range = 0.02-0.06) in all age groups, indicating that females had higher FI scores than males at all ages. This finding was consistent across individual studies. Heterogeneity was high (I^2 = 72.7%), reflecting methodological differences. Meta-analysis of mortality data (13 127 participants) showed that male mortality rates exceeded female mortality rates up until the 90 to 94-years age group. Individual studies reported higher mortality for males at each level of FI, and higher risk of death for males when controlling for age and FI.

Conclusions

The pattern of sex differences in the FI and mortality of older adults was consistent across populations and confirmed a ‘male-female health-survival paradox’.

Keywords: Frailty Index, frailty, sex differences, systematic review, meta-analysis
Sex differences in frailty: A systematic review and meta-analysis

1. Introduction

Within the ageing population, there is significant heterogeneity with respect to health status, care needs and survival. In particular, there is marked discrepancy between the health and survival of the sexes. Despite ‘health advantages’ that contribute to longer life expectancy, females have greater levels of disability, more co-morbidities and poorer self-rated health (Hubbard and Rockwood, 2011). This health differential is present in middle age and, as a result, cannot be solely attributed to longer lifespan (Hubbard and Rockwood, 2011; Mitnitski et al., 2002). Biological, behavioural and social factors have all been identified as potential mediators of the ‘male-female health-survival paradox’ (Hubbard and Rockwood, 2011; Oksuzyan et al., 2008; Verburgge, 1985), but there is still much that needs to be done in order to understand why males and females age in different ways.

In more recent times, ‘frailty’ has emerged as a construct to explore the health status of older adults. ‘Frail’ individuals have poorer health than their age-matched peers and are at an increased risk of adverse outcomes, including disability, institutionalisation and death (Clegg et al., 2013). Therefore, if ‘frailty’ represents the link between health status and poor outcomes in people of the same chronological age, it may serve as a useful paradigm to investigate the male-female health-survival paradox.

Some frailty studies of community-dwelling populations have found that females have higher frailty scores than males (Collard et al., 2012; Mitnitski et al., 2004; Puts et al., 2005). Others have found that females tolerate frailty better, as evidenced through lower mortality rates at any given frailty score (Berges et al., 2009; Mitnitski et al., 2004). However, the male-female health-survival paradox has not been a consistent finding across all frailty studies (Berges et al., 2009; Kulmala et al., 2014; Kulminski et al., 2008; Kulminski et al., 2006; Puts et al., 2005; Saum et al., 2014). The variable results may reflect
differences in the way researchers defined, and therefore, assessed frailty. Frailty scales are generally derived from two key conceptual models of frailty: the phenotypic model (Fried et al., 2001) and the cumulative deficit model (Mitnitski et al., 2002). Whilst there is some overlap in identification of frailty, it is likely that these models capture different groups (Theou et al., 2013). Nevertheless, a recent study demonstrated the male-female health-survival paradox in a sample of older Europeans using seven different frailty scales (derived from the different conceptual models) (Theou et al., 2014). This suggests that the type of frailty scale used to assess participants is unlikely to entirely explain the variability of results. Even so, studies using the same approach to frailty assessment have yielded inconsistent results (Berges et al., 2009; Kulmala et al., 2014), which may indicate subtle differences in the operationalization of a frailty tool. Alternatively, variation across studies may be due to differences in study sample characteristics, such as life-stage (e.g., middle-aged versus old-age) (Saum et al., 2014), disability prevalence (Kulminski et al., 2006), country-of-origin and ethnicity.

Systematic reviews have found the prevalence of frailty to be higher in older females than males, using phenotypic and cumulative deficit models of frailty (Collard et al., 2012; Shamliyan et al., 2013). However, these reviews did not determine whether this sex difference varied across different age groups or whether the degree of frailty differed between the sexes. One systematic review reported hazard ratios linking frailty and mortality, but did not determine whether frailty was more lethal for males or for females (Shamliyan et al., 2013). Furthermore, the majority of studies included in these reviews represented populations from North America, with only small contributions from Europe and Asia.

Currently, there is no consensus as to a standardized tool for frailty measurement. Different tools yield different estimates of frailty prevalence and vary in their content validity, predictive validity and feasibility (de Vries et al., 2011; Theou et al., 2013). A systematic review identified the Frailty Index (FI), representing the cumulative deficit model of frailty, as the most suitable tool for frailty research at the
present time (de Vries et al., 2011). The FI provides a robust, multidimensional summary of health status with strong construct and predictive validity (Rockwood and Mitnitski, 2007, 2011). The characteristics of the FI and its ability to predict all-cause mortality have been confirmed by multiple studies of older cohorts from different cultural backgrounds (Garcia-Gonzalez et al., 2009; Goggins et al., 2005; Mitnitski et al., 2001; Mitnitski et al., 2004; Theou et al., 2013). Since the FI is relatively insensitive to the nature of the variables it contains, it lends itself to meta-analysis. Furthermore, the FI is a continuous variable, which not only provides information about frailty severity but also enables the quantification of health status of people not detected as frail by dichotomous frailty scales. This systematic review aimed to ascertain whether the male-female health-survival paradox is demonstrated when the FI is used to measure frailty. Establishing whether the results of FI studies are consistent in reflecting the sex paradox will help to determine whether the FI model can provide a solid foundation for systematic exploration of biological, psychological and social factors that influence the health trajectories of males and females as they age.

2. Methods

This systematic review and meta-analysis was conducted in accordance with the guidelines established by the Meta-analysis of Observational Studies in Epidemiology (MOOSE) group (Stroup et al., 2008).

2.1. Search Method and Study Selection

The Medline, EMBASE and CINAHL databases were searched for articles up until December 2015. Key words included ‘sex differences’, ‘sexes’, ‘frailty’, ‘frailty index’ and ‘aged’. MeSH headings included ‘sex distribution’, ‘frail elderly’, ‘aged’ and ‘epidemiology’ (see Appendix for full search strategy). After removal of duplicates, assessment involved reviewing the title, the abstract, and finally, the full text. The reference lists of included studies and review articles were hand-searched for additional publications. Experts in the field were contacted via email to identify relevant ‘in press’ articles.
Articles were included if they were written in English, available in full-text and presented observational data from a community-dwelling population, including males and females aged 65 years and older. This age group was chosen in order to maximize the prevalence of frailty and mortality. Studies from Asia with institutionalised as well as community-dwelling participants in the study sample were included in this systematic review. Social factors, such as proximity to children, marital status, financial dependence and access to residential care facilities (particularly in rural areas), were key determinants of institutionalisation in China and Hong Kong around the time that population surveys were conducted (i.e., the early 1990s to 2000s)(Gu et al., 2007; Wan et al., 2008; Woo et al., 1994; Woo et al., 2000). This is in contrast to Western countries, where poor health and functional dependence were likely to be more important factors. Consequently, delineating community-dwelling and institutionalised adults was not thought to be as meaningful in study samples from Asia.

Articles were included if they used the Frailty Index as described by Mitnitski and Rockwood (Mitnitski et al., 2002; Mitnitski et al., 2001; Searle et al., 2008) and reported on the relationship between FI, sex and age as well as the relationship between FI, sex, age and mortality. On the occasion that more than one article reported on the same cohort, the article that presented data of most relevance to the objectives of this systematic review was selected. The inclusion and exclusion criteria are presented in Table One.

< Table One >

After identifying articles meeting selection criteria, studies were then selected for inclusion in the meta-analysis. The aim was to represent the developing world as well as the developed world by collating studies from many different populations. In the event that two studies reported on a similar cohort from the same country or region, the study reporting on the larger, more diverse sample population was
selected. E.G. assessed all identified articles and cross-checked the inclusion and exclusion of full-text articles with N.P. and R.H.

2.2. Data Extraction

Information was collected regarding study design (including setting, follow-up and exclusion criteria); participant characteristics (including age, sex and living circumstances); FI construction (including number and nature of variables and management of missing variables); and statistical methods (including adjustment for potential confounders). Age groups varied between studies (e.g., two, five and ten-year age groups) and FI data were frequently presented in graphical form only. Consequently, E.G. and R.H. contacted study authors to request raw data, including (unweighted) mean FIs (and standard deviations) for males and females in five-year age groups (starting at 65 years) and number of participants per group. In the instance where raw data could not be obtained from study authors, mean FI for each age group was estimated from published graphs.

The included studies presented mortality data in different ways. Where mortality was evaluated using multi-variable regression models, adjusted effect estimates (e.g., hazard ratios) were taken from the text. Mortality rates per age group and FI quartiles were also extracted. Mortality rates (for each age group and sex) were also requested directly from study authors, acknowledging that follow-up periods varied between studies.

2.3. Quality Assessment

Study quality was assessed using a tool based upon Genaidy et al.’s (2007) Epidemiological Assessment Instrument. Domains included study objectives and design, sampling (including eligibility criteria, characteristics of participants and non-participants), outcome measures (particularly FI construction), confounders and co-variates, as well as statistical methods.
2.4. Statistical Analysis

A meta-analysis was conducted to investigate the relationship between sex, age and FI. For this analysis, the mean difference (MD) between female and male mean FI was considered to be the continuous effect measure \( \text{MD} = \text{mean FI}_{\text{Female}} - \text{mean FI}_{\text{Male}} \). The MD for each five-year age group was calculated for included studies. A weighted mean inverse variance (I-V) approach was adopted to estimate the MD for pooled data for each five-year age group (Borenstein et al., 2009). Forest plots were prepared for each age group. Heterogeneity of the effect measure across studies was explored through chi-square testing (Cochrane Q) and the \( I^2 \) statistic. The \( I^2 \) statistic represents the percentage of total variation across studies that is due to heterogeneity rather than chance (Higgins et al., 2003). Death rate for studies with similar follow-up periods were combined for each age group and sex. A summary of mortality results provided by the individual studies was also compiled. Where studies combined the oldest-old participants into one age group (for example, ‘80+’ and ‘90+’ age groups), data were analysed with the 80 to 84-years age group and the 90 to 94-years age group, respectively. The statistical packages R (version 3.1.1) (R Core R Core Team, 2013) and STATA (release 14) (StataCorp, 2015) were used for data analysis.

3. Results

3.1. Search Results

The database search yielded 9144 articles (Figure One). No additional publications were identified through searching reference lists or contacting authors. After removal of duplicates, the titles and abstracts of 6482 articles were assessed, with 6374 identified as not relevant. Subsequently, 108 full-text articles were assessed for eligibility and 99 were excluded. Of the remaining nine relevant articles, two were excluded from further analysis (Lucicesare et al., 2010; Saum et al., 2014) as they reported on samples from two countries (Italy and Germany) that were already represented by a large cross-national
study selected for inclusion in the meta-analysis (Romero-Ortuno and Kenny, 2012). These two studies reported that FI values were higher in older females than older males and males had an increased mortality risk compared with females of the same age and frailty (Lucicesare et al., 2010; Saum et al., 2014). Three studies from China were included in the meta-analysis as samples represented different regions (e.g., mainland China versus Hong Kong) and the combined sample size was comparable with European and North American studies (Goggins et al., 2005; Gu et al., 2009; Shi et al., 2014).

3.2. Study Characteristics

Six articles described longitudinal cohort studies: the Survey of Health, Ageing and Retirement in Europe (SHARE) (Romero-Ortuno and Kenny, 2012), the Chinese Longitudinal Healthy Longevity Study (CLHLS) (Gu et al., 2009), the Mexican Health and Aging Study (MHAS) (Garcia-Gonzalez et al., 2009), the Beijing Longitudinal Study of Aging (BLSA) (Shi et al., 2014), the Irish Longitudinal Study on Ageing (TILDA) (Theou et al., 2015) and a Hong Kong-based population study (Goggins et al., 2005). The seventh study was an integrative analysis of data from the National Population Health Survey, the Canadian Study of Health and Aging, the Australian Longitudinal Study of Ageing, the Sydney Older Persons Studies on Aging, the National Health and Nutrition Examination Survey and the Gothenburg Study (Mitnitski et al., 2005). The characteristics of the studies are presented in Table Two.

Age of participants ranged from 50 years to greater than 105 years. Two studies conducted in Asia included institutionalised adults (Goggins et al., 2005; Gu et al., 2009). In the CLHLS, 4.6% of participants were institutionalised (Gu et al., 2009). The institutionalised proportion of the Hong Kong
study sample was not available for extraction (Goggins et al., 2005). FIs were derived from 40 variables on average (range 33-62 variables), which were mainly self-report items. The authors of the integrative analysis (Mitnitski et al., 2005) were unable to provide raw data given that the study was conducted a decade prior and the results were no longer readily available. Consequently, mean FIs were estimated from their published graph of FI versus age for males and females. The number of participants in each age group could not be extracted. Mean FI values for each age group were obtained for the six remaining studies; however, standard deviation values for the BLSA data were not available (Shi et al., 2014). The mortality follow-up period ranged from two to 10 years (Goggins et al., 2005; Theou et al., 2015) and mortality rates were available for four studies (Garcia-Gonzalez et al., 2009; Goggins et al., 2005; Romero-Ortuno and Kenny, 2012; Shi et al., 2014).

3.3. Quality Assessment

The included studies reported frailty (as measured by FI) and its relationship with mortality for a population of older adults from a geographically defined area. The sampling process was not clearly described in all articles (Mitnitski et al., 2005; Romero-Ortuno and Kenny, 2012; Theou et al., 2015). Potential selection bias was detected in two studies, in which processes may have resulted in under-recruiting older, frailer participants (Garcia-Gonzalez et al., 2009; Goggins et al., 2005). Eligibility criteria (such as place of residence) were not always clearly stipulated (Gu et al., 2009; Shi et al., 2014). In two studies, participants unable to perform assessments independently (Garcia-Gonzalez et al., 2009) or unable to travel for assessment (Theou et al., 2015) were excluded and, as a result, older and frailer participants may have been under-represented. With regards to FI construction, all studies described variables and construction processes clearly. One study included childhood health problems (10 out of 34 variables) in their index (Garcia-Gonzalez et al., 2009), which is at odds with guidelines for constructing an FI (Searle et al., 2008). Missing variables were managed in different ways including imputation of mean values (Gu et al., 2009; Romero-Ortuno and Kenny, 2012), exclusion of variables
from the numerator and denominator (Shi et al., 2014) and exclusion of participants (Garcia-Gonzalez et al., 2009; Mitnitski et al., 2005; Theou et al., 2015). With regards to follow-up data, two studies reported no significant differences in frailty between those followed-up and those lost to follow-up (Goggins et al., 2005; Shi et al., 2014), one study reported no impact on survival analysis (Gu et al., 2009) and four studies did not outline this group of participants in detail (Garcia-Gonzalez et al., 2009; Mitnitski et al., 2005; Romero-Ortuno and Kenny, 2012; Theou et al., 2015). All studies controlled for age and sex when examining the relationship between FI and mortality. One study explored other potential confounders, such as ethnicity, urbanicity and socioeconomic status, but detected only a modest mediating effect (Gu et al., 2009). With respect to the relationship between sex and frailty, potential confounders were not explored. However, this relationship was not the focus of these studies.

3.4. Statistical Analysis

A total of 39 644 participants (54.8% female) aged 65 years and older was collated from six population studies. Of this sample, 7404 participants (18.7% of total) were in the 90 to 94-years age group or older. The integrative analysis by Mitnitski et al. (2005) included 33 851 participants (57.9% female) aged 60 years and older.

3.4.1. FI by age and sex

In each age group, mean FI values varied between studies. For example, in the youngest age group (65 to 69-years) mean FI ranged from 0.09 (CLHLS) to 0.14 (MHAS) for males and 0.10 (CLHLS) to 0.17 (MHAS) for females. In the oldest age group (100+ years), mean FI ranged from 0.21 (SHARE) to 0.32 (CLHLS) for males and 0.38 (CLHLS) to 0.48 (SHARE) for females. Compared with the other studies, the Hong Kong cohort had lower mean FIs and lower rates of deficit accumulation for both sexes (Goggins et al., 2005). The Hong Kong study also found that frailty increased with age until plateauing at 80 to 85 years in males and 90 to 94 years in females (Goggins et al., 2005). In the other studies
(including the integrative analysis study), frailty increased with age in both sexes (Garcia-Gonzalez et al., 2009; Gu et al., 2009; Mitnitski et al., 2005; Romero-Ortuno and Kenny, 2012; Shi et al., 2014; Theou et al., 2015).

A meta-analysis of FI data from five population studies (37,426 participants) was performed. The BLSA study (Shi et al., 2014) and the integrative analysis study (Mitnitski et al., 2005) were not included due to missing data (i.e., standard deviations and sample size, respectively). For all studies, the MD values were positive for each age group (Figure 2). The MD values ranged from 0.01 (95% CI 0.0 – 0.02; CLHLS; Figure 2A) to 0.27 (95% CI -0.03 – 0.57; SHARE; Figure 2H). In all age groups, the pooled MD value was positive (range 0.02-0.06) and the p-value obtained from the test for pooled MD was <0.001. The positive MD values demonstrate that females had higher FI scores than males for any age group.

The overall MD increased with age until the 85 to 89-years age group, at which point it plateaued (Figure 2E). Between-study differences in sample size, mean FI and MD values were more evident in the oldest age groups (Figure 2G-H). Heterogeneity, as measured by the I^2 statistic was 72.7% overall (ranging from zero in the 85 to 89-years age group to 81.8% in the 65 to 69-years age group), indicating methodological differences between studies.

The results of this meta-analysis are consistent with the outcomes of the BLSA study, which found that for any age group, the mean FI was greater in women than in men (p < 0.001 for the difference in the mean) (Shi et al., 2014). Similarly, Mitnitski et al. (2005) showed in their integrative analysis, that at all ages, females, on average, had higher FI scores than males.

3.4.2. Mortality by age and sex
Raw mortality data were obtained for four studies but only two could be analysed together. Mortality rates at approximately two years follow-up were available for the SHARE and MHAS datasets (13,127 participants; Figure 3). In both populations, males demonstrated an increased mortality rate until about 90 years of age. At this point, the female mortality rate exceeded the male mortality rate.

![Figure 3](image)

The BLSA demonstrated that mortality rate increased with age for both sexes, but was higher in males than in females for all age groups (Shi et al., 2014). The CLHLS also found that males had higher mortality rates compared with age-matched females (Gu et al., 2009).

3.4.3. Mortality by FI and sex

Using pooled community-dwelling and institutionalised data (6,427 individuals), Mitnitski et al. (2005) demonstrated that males had an increased mortality rate compared with females at each level of FI. The SHARE and CLHLS also confirmed that mortality increased with FI and males had greater mortality rates despite lower mean FI values (Gu et al., 2009; Romero-Ortuno and Kenny, 2012). The BLSA obtained 95% and 99% FI limits for each age group and sex (i.e., mean FI values of the 5% and 1% frailest individuals) and demonstrated that as individuals approached the 99% frailty limit, mortality rates were higher for males than for females. However, there was little sex difference in mortality rates at the 99% frailty limit (Shi et al., 2014).

3.4.4. Mortality by age, FI and sex

Some studies used multi-variable regression models to evaluate mortality. In the TILDA study, females had reduced risk of mortality (compared with males) when controlling for age and FI (OR 0.539; p < 0.05) (Theou et al., 2015). The MHAS found females had a 33% lower risk of death compared with men.
when controlling for age and frailty (Garcia-Gonzalez et al., 2009). In Hong Kong, multiple regression models found that females had 20% lower risk of death compared with males of the same age and frailty (HR 0.80; p < 0.0005) (Goggins et al., 2005).

4. Discussion

When a Frailty Index is used to evaluate older males and females, there is a consistent relationship between sex, frailty and mortality. This review identified seven large studies of community-dwelling older adults from 19 different countries, across the developed and developing world. Meta-analysis of the data confirmed that, in every age group, females had higher frailty index scores than males. All studies found that females tolerated this frailty better, as demonstrated by a lower mortality rate at any given level of frailty or age. Overall, this systematic review established that sex differences in the FI demonstrate the well-known male-female health-survival paradox.

The difference (MD) between female and male FI scores ranged from 0.02 to 0.06. Given that the average number of FI variables in the included studies was 40, these MD values correspond to between 0.8 and 2.4 additional FI deficits for females when compared with males of the same age group. It is important to consider the clinical relevance of seemingly small differences in the number of FI deficits experienced by males and females. Although the impact of an additional deficit (or deficits) on measures of morbidity (such as, self-rated health or disability) has not been quantified in the literature to date, the clinical relevance of increasing values of the FI with respect to mortality have been reported. For example, in the BLSA, a 0.1 point increase in FI (corresponding to 3.5 deficits) increased the risk of eight-year mortality by 13% in models adjusted for age and sex (Shi et al., 2011). Similarly, in the TILDA and Canadian Study of Health and Aging studies, a 0.01 point increase in FI (corresponding to approximately one third of a deficit in each study) was associated with a 3.6% and a 2% increase in risk of two-year and six-year mortality, respectively (Howlett et al., 2014; Theou et al., 2015). Even though
the relationship between FI and mortality could not be quantified in this meta-analysis, the results of these other studies indicate that even small changes in the FI score (corresponding to less than one FI deficit) influence risk of death. An MD value as small as 0.02 is an important finding because it emphasizes that females, with their (marginally) higher FI scores, should face a higher risk of mortality than males. The fact that they do not is the sex paradox.

This meta-analysis also showed that the gap between male and female FI scores increased with age until the 85 to 89-years age group, at which point it plateaued. A widening of the gap between sexes may have been consistent with a survival effect. That is, the difference between male and female frailty increased because the frailest males died, leaving the healthiest males in the sample. However, a survival effect would be expected to continue into the oldest-old age groups. In this meta-analysis, those over the age of 80 years represented between 2.9% and 70% of the samples (Gu et al., 2009; Theou et al., 2015) and, as a result, smaller sample size may have impacted results in the oldest age groups. Mean FI values in these age groups varied markedly between studies, reflecting differences in recruitment processes, sample size and study eligibility criteria. Given the large contribution of the CLHLS data to the oldest age groups in this meta-analysis (versus a smaller contribution by SHARE), the impact of ethnicity and cultural differences must also be considered (Gu et al., 2009). Despite methodological issues, females in these older age groups consistently scored higher mean FI values. This finding is in keeping with the literature, which identifies greater co-morbidity, geriatric syndromes and disability (self-reported and performance-based) in the oldest old females when compared with age-matched males (Hazra et al., 2015; Nybo et al., 2001). Beyond the survival effect, it is possible that there are sex differences specific to the oldest old population, such as different patterns and rates of deficit accumulation.
Although the pattern of sex differences in frailty and mortality were consistent between studies, mean FI values (per age group) did vary. Inherent differences between sample populations with respect to ethnicity, gender roles, social characteristics and health systems, likely account for a significant proportion of the heterogeneity between these studies. Selection and exclusion bias may have also contributed to the observed differences in frailty. For example, excluding individuals that were unable to travel for assessment or participate in an interview likely underestimated the frailty of some populations (Garcia-Gonzalez et al., 2009; Theou et al., 2015). Similarly, in some studies, selection processes may have favoured the inclusion of less frail individuals (Garcia-Gonzalez et al., 2009; Goggins et al., 2005).

The plateau of mean FI in older age groups seen only in the Hong Kong study (Goggins et al., 2005) may indicate that respondents were more robust than non-respondents, a pattern detected in other populations (McCaul et al., 2015). The impact of including institutionalised participants in some studies is not known and may have contributed to heterogeneity. In general, the FIs of included studies used a similar number of variables that were mainly self-report and in keeping with recommended guidelines. Although there were small between-study differences in FI construction, it is generally accepted that characteristics of the FI are insensitive to the number and nature of variables used to create an index (Rockwood and Mitnitski, 2011).

The sex difference in FI values demonstrated by this systematic review was also detected in American community-dwelling adults, even after adjusting for differences in social roles, socioeconomic status, lifestyle and psychosocial stressors (Cigolle et al., 2009; Yang and Lee, 2010). In another US study, FI frequency distributions showed that surviving males were ‘healthier’ than surviving females, but the FI appeared to be sex-insensitive in the younger and older age groups (Kulminski et al., 2007a). This study only included community-dwelling adults with disability and, as a result, this finding may indicate that sex differences in frailty vary in specific populations. Nevertheless, this study still found that males faced a higher risk of death than females at any given level of frailty (Kulminski et al., 2007b).
A frequent hypothesis for the sex paradox is that females have poorer health status because they live longer. That is, females are likely to acquire more deficits over time and they are likely to have these deficits for longer. However, these results show that at any given age, females have higher frailty scores than males. This suggests that there is an inherent sex difference in the prevalence of deficits. The SHARE group demonstrated in their study of people over 50 years of age that males and females both acquired new co-morbidities with age but females acquired slightly more deficits overall (Avedano and Mackenbach, 2008). The higher female prevalence of ‘non-lethal’ diseases that impact negatively on function and quality of life has been identified in many studies (Case and Paxson, 2005; Crimmins et al., 2010; Ferrucci et al., 2003; Gorman and Read, 2006; Shi et al., 2014), as have the higher rates of self-reported disability in females (Avedano and Mackenbach, 2008; Crimmins et al., 2010; Gorman and Read, 2006; Merrill et al., 1997; Shi et al., 2014). Importantly, self-reported disability seems to correlate well with performance, with females testing as more disabled than males on performance measures (Merrill et al., 1997). Sex differences in biological factors, such as inflammatory cytokines, sarcopenia, abdominal adiposity and cognitive impairment, are emerging in the literature and probably underpin sex differences in co-morbidity and disability (Canon and Crimmins, 2011; Hubbard et al., 2010). Psychosocial factors, such as healthcare utilisation and self-report behaviours have also been implicated in the sex differential (Oksuzyan et al., 2008).

Similar biological and psychosocial factors contribute to sex differences in mortality. In particular, risk-related activities, access to preventative medicine and prevalence of ‘lethal’ co-morbidities have been identified as contributing to the male-female mortality gap (Case and Paxson, 2005; Gorman and Read, 2006; Hubbard and Rockwood, 2011; Oksuzyan et al., 2008). But, these factors do not explain why females tolerate deficit accumulation better than males (as demonstrated by the lower female mortality rate at any level of FI). One hypothesis is that the sex paradox transpires because males have lower
physiological reserve than females. In the BLSA study of FI limits, the 99% limit was lower for males (mean 0.44, max 0.61) than females (mean 0.52, max 0.69) and males reached the limit at a younger age (Shi et al., 2014). Therefore, whilst all individuals at the 99% FI limit faced a very high risk of mortality, males faced it earlier. Similarly in the MHAS, males demonstrated an increased risk of death at an FI above 0.21, whereas females risk of death emerged at an FI of greater than 0.35 (Garcia-Gonzalez et al., 2009). Together, these studies support the hypothesis that males and females have different amounts of physiological reserve.

Another hypothesis is that deficits influence the risk of mortality in different ways for each sex. Some deficits (such as, arthritis, hearing aids and thyroid disease) appear to be associated with a lower risk of five-year mortality than other deficits (Shi et al., 2014). A central tenet of the cumulative deficit model is that the number of, rather than the nature of, deficits is critical to frailty and adverse outcomes. Hence, these preliminary findings highlight the need for further investigation, particularly with regards to potential sex differences.

It is also possible that investigations of the male-female health-survival paradox to date have not captured important social factors that impact the sexes in different ways. For example, an individual’s wealth and the level of neighbourhood deprivation impacts on frailty and mortality but the size and direction of the impact may differ between the sexes (Gu et al., 2009; Lang et al., 2009; Major et al., 2010; Romero-Ortuno, 2014; Woo et al., 2005). Factors such as marital status, education level and social supports have all been reported to impact on male and female frailty (Andrew et al., 2008; Gorman and Read, 2006; Gu et al., 2009; Szanton et al., 2010; Woo et al., 2005).

Recently, studies examining test-based and laboratory-based FIs have revealed new patterns of sex differences (Howlett et al., 2014; Theou et al., 2015). ‘Subclinical’ deficit accumulation may be
measurable and be relevant in predicting adverse outcomes (Mitnitski et al., 2015), particularly in men where clinical frailty may be low but the risk of adverse outcome is high. This new research indicates the cumulative deficit model provides further avenues to explore sex differences in frailty.

This systematic review identified several large population-based studies from developing and developed countries and performed a meta-analysis of FI from over 37,000 adults. However, this review is not without its limitations. The data, derived from a relatively small number of studies, were heterogeneous, with variability in mean FI values across studies. Despite this, there was remarkable consistency in the pattern of sex differences determined by all included studies. Due to the differences in study design, mortality data were unable to be obtained in a consistent format and only a limited meta-analysis could be performed for this data. Data were estimated from the published graphs of Mitnitski et al.’s (2005) integrative analysis and the accuracy of this process is not known.

5. Conclusion
In summary, this systematic review and meta-analysis has shown that sex differences in the FI and mortality reflect the long-described male-female health-survival paradox. This result establishes the FI model as a solid foundation for the investigation of sex differences in frailty. Future research should identify health deficits and assets that impact the health status and survival of males and females in different ways. Hypotheses regarding sex differences in physiological reserve and ‘subclinical frailty’ should also be explored.

Acknowledgements
The authors of this systematic review gratefully acknowledge the authors of the included studies for their assistance with the provision of raw data for the meta-analysis.
This paper uses data from SHARE Waves 1 and 2 (release 2.5.0, as of 24th May 2011). The SHARE data collection has been primarily funded by the European Commission through FP5 (QLK6-CT-2001-00360), FP6 (SHARE-I3: RII-CT-2006-062193, COMPARE: CIT5-CT-2005-028857, SHARELIFE: CIT4-CT-2006-028812) and FP7 (SHARE-PREP: N°211909, SHARE-LEAP: N°227822, SHARE M4: N°261982). Additional funding from the German Ministry of Education and Research, the U.S. National Institute on Aging (U01_AG09740-13S2, P01_AG005842, P01_AG08291, P30_AG12815, R21_AG025169, Y1-AG-4553-01, IAG_BSR06-11, OGHA_04-064) and from various national funding sources is gratefully acknowledged (see www.share-project.org for full list of funding institutions).

Conflict of Interests

The authors report no conflict of interests.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.
References


StataCorp, 2015. Stata statistical software: Release 14. StataCorp LP, College Station, TX.


Table 1: Exclusion and inclusion criteria for systematic review.

<table>
<thead>
<tr>
<th>Article Selection Criteria</th>
<th>Included</th>
<th>Excluded</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Publication Criteria</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Published and ‘in press’ articles in peer reviewed scientific journals</td>
<td></td>
<td>• Reviews, book chapters, editorials, dissertations, theses and conference abstracts</td>
</tr>
<tr>
<td>• Reporting original research results</td>
<td></td>
<td>• Full text not available</td>
</tr>
<tr>
<td>• Written in English</td>
<td></td>
<td>• Main body of article written in a language other than English</td>
</tr>
<tr>
<td><strong>Study Design</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Observational data (cohort, cross-sectional, longitudinal studies)</td>
<td></td>
<td>• Qualitative studies</td>
</tr>
<tr>
<td>• Quantitative studies</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Study Population</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Sample included adults &gt;65 years</td>
<td></td>
<td>• Mean age at baseline ≤60 years</td>
</tr>
<tr>
<td>• Community dwelling</td>
<td></td>
<td>• Male- or female-only samples</td>
</tr>
<tr>
<td>• Studies of specific populations (e.g., disease-specific, non-disabled, cognitively-intact)</td>
<td></td>
<td>• Duplication of study cohort (when an article using that study population has already been selected for inclusion in the systematic review)</td>
</tr>
<tr>
<td><strong>Outcome Measures</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Frailty Index</td>
<td></td>
<td>• Frailty Index constructed from laboratory values and performance measures only</td>
</tr>
<tr>
<td>• Minimum of 30 variables (majority self-report)</td>
<td></td>
<td>• Composite adverse outcome measures only (e.g., falls + hospitalization + mortality)</td>
</tr>
<tr>
<td>• Variables include signs, symptoms, diseases and disabilities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Variables represent different domains</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Mortality</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Independent Variables</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Study Details</td>
<td>Characteristics of Study Population</td>
</tr>
<tr>
<td>-------</td>
<td>---------------</td>
<td>-------------------------------------</td>
</tr>
<tr>
<td>Romero-Ortuno and Kenny (2012)</td>
<td>Survey of Health, Ageing and Retirement in Europe (SHARE) 12 European countries 2004/05 – 2005/06 Longitudinal cohort study</td>
<td>• 29 905 participants (54.2% female)  • Community-dwelling adults 50-104yrs  • Mean age: 64.8yrs female; 64.3yrs male  • 9% participants &gt; 80yrs  • Mortality data for 19 789 participants  • Follow-up at 2.4 years (average)</td>
</tr>
<tr>
<td>Gu et al. (2009)</td>
<td>Chinese Longitudinal Healthy Longevity Study (CLHLS) Random selection of cities from 22/31 provinces in China 2002 - 2005 Longitudinal cohort study</td>
<td>• 15 919 participants (56.6% female)  • Community-dwelling and institutionalised adults 65-109yrs  • Mean age: ~85yrs female; ~82yrs males  • 70% participants &gt; 80yrs  • Mortality data for 13 861 participants  • Follow-up at 3 years</td>
</tr>
<tr>
<td>Goggins et al. (2005)</td>
<td>Hong Kong 1990/91 – 2000/01 Longitudinal cohort study</td>
<td>• 2032 participants (50.8% female)  • Community-dwelling and institutionalised adults 70-107yrs  • Mean age: 79.3yrs females; 80.1yrs males  • 42% participants &gt; 80yrs  • Mortality data for 1382 participants  • Follow-up at 3 years and 5 years</td>
</tr>
<tr>
<td>Garcia-Gonzalez et al. (2009)</td>
<td>Mexican Health and Aging Study (MHAS) Mexico 2001 – 2003 Longitudinal cohort study</td>
<td>• 4082 participants (52.5% female)  • Community-dwelling adults 65-105yrs  • Mean age: 73yrs females and males  • 15% participants &gt;80yrs  • Mortality data for 4082 participants  • Follow-up at 2 years</td>
</tr>
<tr>
<td>Shi et al. (2014)</td>
<td>Beijing Longitudinal Study of Aging (BLSA) Beijing 1992 – 1997</td>
<td>• 3257 participants (50.9% female)  • Community-dwelling adults 55-85yrs  • Mean age: females 75.3yrs; males 74.6yrs  • 18% participants &gt;80yrs  • Mortality data for 2983 (91.6%)</td>
</tr>
<tr>
<td>Study Type</td>
<td>Study Name</td>
<td>Data Collection Period</td>
</tr>
<tr>
<td>----------------------------</td>
<td>-------------------------------------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Longitudinal cohort study</td>
<td>Irish Longitudinal study on Ageing (TILDA)</td>
<td>Ireland 2010-2012</td>
</tr>
<tr>
<td></td>
<td>Theou et al. (2015)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cross-sectional study</td>
<td>National Population Health Survey (NPHS)</td>
<td>Canada 1996</td>
</tr>
<tr>
<td></td>
<td>Canadian Study of Health and Ageing (CSHA)</td>
<td>Canada 1991-1996</td>
</tr>
<tr>
<td></td>
<td>CSHA – Exam</td>
<td>Canada 1996-2002</td>
</tr>
<tr>
<td></td>
<td>Australian Longitudinal Study of Ageing (ALSA)</td>
<td>Australia 1992</td>
</tr>
<tr>
<td></td>
<td>Sydney Older Persons Studies on Ageing (SOPSA)</td>
<td>Australia 1992-1996</td>
</tr>
<tr>
<td>Cross-sectional study with mortality follow-up</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
</tbody>
</table>
| Gothenburg Study (H-70) Sweden 1971-1996 Birth cohort study | • 965 participants  
• Community-dwelling adults 70yrs | • 40-item FI  
• 36 self-report items; 4 measured items (cholesterol, blood glucose, BMI >30, BMI <20) |  |

BMI – body mass index; MMSE – mini mental status examination; SBP/DBP – systolic/diastolic blood pressure; GDS – Geriatric Depression Scale  
RR – relative risk; HR – hazard ratio; OR – odds ratio; CI – 95% confidence interval
Appendix

Search Strategy

Limits used with each database – human, English

Pubmed

(MH "Sex") OR (MH "Sex Distribution") OR (MH "Sex Characteristics") OR "sex differences" OR "gender" OR "male" OR "female" OR "sexes"
AND
(MH "Frail Elderly") OR "frail" OR "frailty" OR "frailty index"
AND
(MH "Aged") OR "aged" OR "elderly" OR "old" OR (MH "Geriatrics") OR "geriatric"
AND
(MH "Epidemiology") OR "epidemiology" OR (MH "Population Characteristics") OR "population"

Cinahl

(MH "Sex") OR (MH "Sex Distribution") OR (MH "Sex Characteristics") OR "sex differences" OR "gender" OR "male" OR "female" OR "sexes"
AND
(MH "Frail Elderly") OR "frail" OR "frailty" OR "frailty index"
AND
(MH "Aged") OR "aged" OR "elderly" OR "old" OR (MH "Geriatrics") OR "geriatric"

Embase

'aged'/exp OR 'geriatrics'/exp OR 'old' OR 'aged' OR 'elderly' OR 'geriatric'
AND
'frail elderly'/exp OR 'frailty' OR 'frail' OR 'frailty index'
AND
'sex'/exp OR 'sex ratio'/exp OR 'sex differences' OR 'gender and sex'/exp OR 'male'/exp OR 'female'/exp OR 'sexes'
AND
'epidemiology'/exp OR 'population and population related phenomena'/exp
Figures

Figure 1: Flow diagram of study selection.

- Database Search: 9144 articles
- 2662 duplicates excluded
- 6482 articles
- 6374 articles excluded on title and abstract
- 108 articles
- 99 articles excluded on full text
  - FI not measured (n=35)
  - No mortality outcome or composite measure only (n=15)
  - Specific population (n=12)
  - Type of article (review, not English) (n=6)
  - Duplication of study cohort (n=28)
  - Outcomes not stratified by sex (n=3)
- 9 relevant articles
- 7 articles included in meta-analysis
Figure 2 A-H: Meta-analysis forest plots for each five-year age group demonstrating the mean difference between female and male FI scores.

(W) MD = (weighted) mean difference; CI = confidence interval; I-squared = heterogeneity measure (I²)

HK (Hong Kong Study) (Goggins et al., 2005); TILDA (Theou et al., 2015); SHARE (Romero-Ortuno and Kenny, 2012); CLHLS (Gu et al., 2009); MHAS (Garcia-Gonzalez et al., 2009)

Note:
Boxes mark the point estimate of the MD for each study with bars representing 95% confidence intervals. The size of the box denotes the contribution of each study to the meta-analysis (the study weight). A diamond represents the pooled data, with the middle marking the estimated MD (the summary effect measure) and the left and right points representing the corresponding confidence interval. Since MD = FI(Female) – FI(Male), an MD value greater than zero indicates that the female FI score is higher than the male FI score.

A 65 to 69-years

B 70 to 74-years
G 95 to 99-years

H 100+ years
Figure 3: Combined mortality rates in five-year age groups for SHARE and MHAS studies (follow-up at approximately two years). SHARE (Romero-Ortuno and Kenny, 2012); MHAS (Garcia-Gonzalez et al., 2009)
Highlights

- The male-female health-survival paradox: females live longer than males but with poorer health.

- Population studies using the Frailty Index to evaluate older adults consistently show that at any given age, females have higher FI scores than males and females tolerate this frailty better, as demonstrated by a lower mortality rate at any given level of frailty or age.

- The Frailty Index model provides a framework for investigating the mechanisms for this sex paradox, including differences in co-morbidity, disability and social determinants of health.