Frailty Index as a Predictor of Mortality: A Systematic Review and Meta-analysis

Gotaro Kojima, MD¹; Steve Iliffe, FRCGP¹; Kate Walters, PhD¹.

¹ Department of Primary Care and Population Health, University College London, London, UK

Corresponding Author:
Gotaro Kojima, MD
Department of Primary Care and Population Health
University College London (Royal Free Campus)
Rowland Hill Street,
London, NW3 2PF, UK
Phone: +44 (0)20 7794 0500
Fax: +44 (0)20 7472 6871
Email: gotarokojima@yahoo.co.jp
ABSTRACT

BACKGROUND
Two popular operational definitions of frailty, the frailty phenotype and Frailty Index (FI), are based on different theories. Although FI was shown to be superior in predicting mortality to the frailty phenotype, no meta-analysis on mortality risk according to FI has been found in the literature.

METHODS
An electronic systematic literature search was conducted in August 2016 using four databases (Embase, Medline, CINAHL and PsycINFO) for prospective cohort studies published in 2000 or later, examining the mortality risk according to frailty measured by FI. A meta-analysis was performed to synthesise pooled mortality risk estimates.

RESULTS
Of 2,617 studies identified by the systematic review, 18 cohorts from 19 studies were included. Thirteen cohorts showed hazard ratios (HRs) per 0.01 increase in FI, six cohorts showed HRs per 0.1 increase in FI and two cohorts each showed odds ratios (ORs) per 0.01 and 0.1 increase in FI, respectively. All meta-analyses suggested that higher FI was significantly associated with higher mortality risk (pooled HR per 0.01 FI increase=1.039, 95%CI=1.033-1.044, p<0.001; pooled HR per 0.1 FI increase=1.282, 95%CI=1.258-1.307, p<0.001; pooled OR per 0.01 FI increase=1.054, 95%CI=1.040-1.068, p<0.001; pooled OR per 0.1 FI increase=1.706, 95%CI=1.547-1.881, p<0.001). Meta-regression analysis among 13 cohorts with HR per 0.01 increase in FI showed that the studies with shorter follow-up periods and with lower female proportion were associated with higher mortality risks by FI.

CONCLUSIONS
This systematic review and meta-analysis was the first to quantitatively demonstrate that frailty measured by the FI is a significant predictor of mortality.
INTRODUCTION
Frailty has been gaining increasing scientific attention over the last few decades. Frailty is generally considered to be a state characterised by reduced physiological reserve and loss of resistance to stressors caused by accumulated age-related deficits.[1] It has been shown that those who are frail are predisposed to various negative health outcomes, such as falls, fractures, hospitalisation, nursing home placement, disability, poor quality of life and dementia.[2-8]

Two of the most popular operational definitions of frailty are the frailty phenotype by Fried and colleagues, using data from the Cardiovascular Health Study,[9] and the Frailty Index (FI) by Rockwood, Mitnitski and colleagues, using the Canadian Study of Health and Aging (CSHA).[10] These two approaches are based on different theories.[11] Frailty phenotype describes frailty as a biological syndrome with specific phenotypic presentations and defines frailty as having three or more of five physical components: unintentional weight loss; self-reported exhaustion; weakness; slow walking speed; and low physical activity.[9] The frailty phenotype is a well-validated and the most frequently used measure in research and clinical practice. On the other hand, this definition has been criticised for being quite narrow in focus, and for not including potentially important components of frailty such as cognitive impairment.[1, 12, 13] By contrast, the concept of the FI is that frailty is a state caused by the accumulation of health deficits during the life course and that the more deficits one has, the more likely one is to be frail.[10] The FI is calculated as a ratio of the number of deficits present to the number of total deficits considered.[10] The deficits can be symptoms, signs, diseases, disabilities, laboratory, radiographic, or electrocardiographic abnormalities and social characteristics.[14] While the exact operationalisation of the FI has varied between studies, standard criteria for constructing a FI are used.[14] Frailty is a strong predictor of mortality,[1] as has been shown by previous systematic reviews.[15-17] Two of these reviews systematically collected studies that used different frailty definitions, including frailty phenotype and the FI, and demonstrated that frailty consistently increased the risk of death in most studies.[15, 16] These reviews just listed mortality risk estimates per different units of the FI from the original papers, therefore it is not possible to directly compare these estimates and no meta-analysis was conducted.[15, 16] The third paper conducted a meta-analysis using the data from only studies using frailty phenotype and showed frailty and pre-frailty significantly predicted mortality in a graded manner.[17] Although the FI was shown to be superior in predicting mortality and other health outcome risks to frailty phenotype in a head-to-head comparison,[18, 19] to the best of our knowledge, no meta-analysis on mortality risk according to the FI has been found in the literature. This may be partially because the previous studies provided mortality risks according to different units of the FI, such as per 0.01 of the FI, 0.1 of the FI or per additional deficit, or according to frailty groups based on arbitrary cutpoints of the FI. Therefore, the objectives of this study are as follows: (1) to conduct a systematic search of the literature for prospective studies examining mortality risk according to frailty defined by the FI; and (2) to combine the effect sizes to synthesise pooled risk estimates of mortality by standard units of the FI, per 0.01 or 0.1 of the FI’s increment.

METHOD
Data source and search strategy
An electronic systematic literature search was conducted in August 2016 by a clinician researcher (GK) based on a protocol developed according to the PRISMA statements.[20] Embase, Medline, CINAHL Plus and PsycINFO were searched for studies published in 2000, given that the first FI paper was published in 2001,[10] or later using a combination of
Medical Subject Heading (MeSH) and text terms without language restriction. The search terms used were ("Mortality (MeSH)" OR "Death (MeSH)" OR "Death and Dying" OR “mortality” OR “death*”) AND ("Rockwood K (as author)” OR “Mitnitski A (as author)” OR “Rockwood” OR “Mitnitski” OR “frailty index” OR “FI”). The names of Professors Rockwood and Mitnitski were used as a search term as they developed the FI and have since published multiple papers using the FI. We also repeated the literature search in July 2017 using “accumulated deficit*”, “cumulative deficit*” and “deficit accumulation” along with abovementioned mortality related terms for additional studies. References of the relevant articles and reviews were also reviewed for additional studies. Forward citation tracking was also conducted on Google Scholar website for the three previous review papers.[15-17]

Eligibility criteria
The following inclusion and exclusion criteria were used.
Inclusion criteria:
1. Prospective study design
2. Adult population with mean age of 20 or greater
3. More than half of the cohort in the community (CSHA included approximately 10% of institutionalised people[10])
4. Baseline frailty defined by the FI constructed according to the published standard methodology[14]
5. Subsequent all-cause mortality risk assessed as hazard ratio (HR) or odds ratio (OR) per 0.01 or 0.1 increase in FI
Exclusion criteria:
1. Selected populations, such as ones with a certain disease or medical condition
2. Mortality risk per additional deficit or per worsening of frailty subgroups, such as by tertile or arbitrary cut-points.
3. Conference presentations, review articles, editorials, comments, or dissertations.

Study selection
The studies identified by the systematic review were assessed using the above inclusion and exclusion criteria by one author (GK). Initially the titles and abstracts were reviewed, and full texts were retrieved for articles that were considered to be eligible or to need a further assessment for eligibility. The full texts and reference lists were examined to identify potentially eligible studies. The original authors were contacted for clarification, if needed. If multiple studies showed the same effect measures using the same cohort, or one study provided multiple results with different conditions, such as for different follow-up periods, the results with the larger number of cohorts, the larger number of deficits used to construct the FI, or longer durations were selected. Each cohort only contributed data once per meta-analysis.

Data extraction
Data extracted from the included studies by the author (GK), using a standardised form, were first author, study name if any, publication year, location, population characteristic, sample size, proportion of female participants, mean age, age range, number of deficits used to create the FI and follow-up period. HRs or ORs of all-cause mortality per 0.01 or 0.1 increase in the FI along with 95% confidence interval (CI) were also collected. The effect measures adjusted confounders were preferred over crude ones.

See Appendix 1 for methodological quality assessment and statistical analysis.
RESULTS

Selection processes

The systematic search of the literature using four electronic databases (Embase, MEDLINE, CINAHL Plus and PsycINFO) yielded 2,611 studies. Six additional studies were found by other source. Of the 2,617 studies, 651 duplicate studies were excluded. The title and abstract screening further excluded 1,891 studies, leaving 75 studies. Full-text review of these 75 studies excluded 56, due to the following reasons: no HR or OR for mortality provided (n=25); effect measures per change in frailty groups based on the FI (n=17); effect measures per each additional deficit (n=4); non-standard FI used (n=3); the same cohort used (n=3); selected population (hospitalized patients) (n=1); unit of the FI for effect measures not clearly documented (n=3). Among these excluded studies, the findings of 28 studies providing mortality risks as HR or OR by frailty status based on the FI in general adult populations were summarised in Appendix 2. All the studies consistently showed worse frailty status defined by the FI in various ways, such as per deficit or grouping, was significantly associated with higher mortality risks.

Nineteen studies were left (the references are listed in Appendix 3) and assessed for methodological quality using the modified 8-item Newcastle-Ottawa scale. All studies met five or more of the eight items and were considered to have adequate methodological quality (range=5-7, mean=6.1).

Two studies provided HR per 0.01 increase in the FI using the Survey of Health, Ageing and Retirement in Europe (SHARE).[23, 24] The study with the larger number (n=37,546) showed that all of adjusted hazard ratio and upper and lower limits of 95% CI were the same at 1.04 (aHR=1.04, 95%CI=1.04-1.04),[23] which was not possible to be included in the meta-analysis. Therefore, the other study (n=36,306) was used instead (aHR=1.05, 95%CI=1.05-1.06).[24] A study showed 2-year, 4-year and 7-year mortality risks (age- and gender-adjusted HRs=1.04 (95%CI=1.03-1.04), 1.03 (95%CI=1.03-1.04) and 1.03 (95%CI=1.03-1.03), respectively).[25] Since the 7-year mortality HR could not be used for the same reason above, the 4-year mortality HR was used for the meta-analysis instead. One study was included after confirmation with the study authors regarding a FI unit used to calculate the effect measures (HR per 0.1 increase in the FI).[26] Additional data (HR per 0.01 increase in the FI) were also provided by the authors of this study[26] and included in the meta-analysis. Four series of meta-analyses were conducted for HR per 0.01 increase in the FI (n=12), HR per 0.1 increase in the FI (n=4), OR per 0.01 increase in the FI (n=2) and OR per 0.1 increase in the FI (n=2). A flow chart of the systematic literature review is shown in Figure 1.

Characteristics of selected studies

Table 1 presents characteristics and outcomes of the included studies. A total of 18 cohorts were used by 19 studies, which were summarised according to unit of the FI used to calculate effect measures (HR per 0.01 of the FI, HR per 0.1 of the FI, OR per 0.01 of the FI, OR per 0.1 of the FI). Four cohorts from Canada were used by six studies,[23, 27-31] three cohorts from the UK were used by two studies,[32, 33] four cohorts from the US were used by four studies,[14, 18, 34, 35] four cohorts from China were used by three studies,[25, 26, 36] two cohorts, both of which consisted of multinational European populations, were used by three studies[24, 37, 38] and lastly one Dutch cohort was used by one study.[39] The sample sizes ranged from 754[14] to 36,306[24]. Two female only cohorts were used by three studies[28, 29, 32] and two male only cohorts were used by three studies.[35, 37, 38] The remaining cohorts were mixed with approximately 50-70% women. The number of deficits used to create the FI ranged from 23[23] to 70.[24, 31] The follow-up periods varied with the shortest
of 2 years[24, 33] and the longest of 19 years.[39] Twelve studies provided HR for mortality risk per 0.01 increase in the FI for 13 cohorts,[14, 18, 23, 25-27, 30, 31, 34, 36, 37, 39] four studies provided HR per 0.1 increase in the FI for six cohorts,[26, 31, 32, 38] two studies provided OR per 0.01 increase in the FI for two cohorts,[28, 33] and two studies provided OR per 0.1 increase in the FI for two cohorts.[29, 35] All included studies provided effect measures adjusted for at least age and gender, or age only in male only or female only cohorts, except for one study[18] providing an unadjusted effect measure.

**Frailty Index as a predictor of mortality**

**Meta-analysis of studies using HR**

HRs of mortality per 0.01 increase in the FI from the 13 cohorts were combined using a random-effects model due to the significant heterogeneity (p<0.001, \( I^2 = 86\% \)). Frailty was a significant predictor of mortality (13 cohorts: pooled HR=1.039, 95%CI=1.033-1.044, p<0.001). Combining HRs per 0.1 increase in the FI from six cohorts using a fixed-effect model (heterogeneity p=0.11, \( I^2 = 45\% \)) also showed that frailty significantly predicted mortality (6 cohorts: pooled HR=1.282, 95%CI=1.258-1.307, p<0.001). (Figure 2 A B)

**Meta-analysis of studies using OR**

Four studies provided OR as a risk measure of mortality. Two studies showed ORs per 0.01 increase in the FI[28, 33] and another two studies showed ORs per 0.1 increase in the FI.[29, 35] fixed-effects models were used (heterogeneity p=0.23 and 0.24, \( I^2 = 30\% \) and 29%, respectively) and both showed that frailty is a significant predictor of mortality (2 cohorts: pooled OR per 0.01 increase in the FI=1.054, 95%CI=1.040-1.068, p<0.001; 2 cohorts: pooled OR per 0.1 increase in the FI=1.706, 95%CI=1.547-1.881, p<0.001, respectively). (Appendix 4 A B)

See Appendix 1 for meta-regression and subgroup analysis and publication bias assessment.
DISCUSSION
The current study identified 19 studies that longitudinally examined mortality risk according
to frailty measured by the FI in 18 cohorts and provided the effect measured as HR or OR per
0.01 or 0.1 increase in the FI. The meta-analysis quantitatively combined mortality risks
based on frailty measured by the FI and consistently showed increased mortality risk
according to the FI regardless of different types of the effect sizes and per units of the FI..
Although the included studies constructed the FI based on different numbers and types
of deficits, in addition to various populations and study settings, it is of note that the effect
measures were in relatively narrow ranges and may support the robustness of this
accumulation deficit frailty model.

Although in general age is a strong predictor of mortality, the mean age of the cohorts was
not a significant modulator in the association between the FI and mortality in the meta-
regression analysis. Furthermore, subgroup analysis also showed that pooled estimates of
studies with a mean age of >65[14, 18, 24-26, 30, 36] and <65[23, 27, 37] (mostly middle
aged with the mean age ranging from 44 to 60.2) were almost identical (8 cohorts: pooled
HR=1.04, 95%CI=1.03-1.05, p<0.001, I²=84%, 3 cohorts: pooled HR=1.05, 95%CI=1.03-
1.07, p<0.001, I²=92%, respectively). This suggests the FI is a good indicator of mortality
risk not only among older people but also among younger populations, regardless of age.

Two study characteristics were found in the meta-regression analysis to be related to the
association between frailty and mortality: follow-up period and female proportion. In general,
women live longer but have more disabilities than men, known as the male-female health-
survival paradox.[40] Given the FI can be regarded as a measure of biological age[10] and
prevalence of frailty is higher among women than men,[9] it is to be expected that female
gender is associated with lower mortality risk according to frailty in the meta-regression
analysis. Regarding the follow-up period, the meta-regression analysis suggests shorter
follow-up periods are associated with higher mortality risk according to the FI. Frailty is a
dynamic state and known to change over time, mostly worsening rather than improving.[41]
The longer follow-up periods imply that as participants get older they usually get frailer. This
may be why the reason the association between frailty and mortality became less prominent
in studies with longer follow-up periods. The studies using the same cohorts with different
lengths of follow-up showed overall comparable results with little difference.[14, 23, 24] In
SHARE, 2-year mortality (aHR=1.05)[24] was slightly higher than 5-year mortality
(aHR=1.04),[23] while 9-year mortality (aHR=1.03)[14] was slightly lower than 12-year
mortality (aHR=1.04)[23] in the Yale Precipitating Events Project.

This study’s findings should be interpreted with caution due to some limitations. First, all
processes of the systematic review and meta-analysis were conducted by one investigator.
Second, during the study selection, a large number of studies that used the FI to examine
mortality risk were excluded because they did not provide HR or OR for mortality (n=25);
the effect measures provided were based on frailty groups defined by different cut-off points
(n=17); or on each additional deficit (n=4). Although not all, at least some of them could
potentially have been included in the meta-analysis. Lastly, the effect measures and upper and
lower limits of 95% CI in many of the included papers were rounded to two decimal places,
which could potentially lead to a miscalculation of standard error or weighting in the meta-
analysis, especially when effect measures were calculated per 0.01 increase in the FI and
were therefore relatively smaller.

The current study has multiple strengths. The search strategy of the systematic review of the
literature was robust and reproducible, using comprehensive search terms in multiple electronic databases. Additional data were also acquired from the original study’s authors.[26] The included studies were also assessed for heterogeneity, methodological quality, and publication bias, and a high degree of heterogeneity was further explored by meta-regression analysis and subgroup analysis. The data from included studies were based on a FI constructed according to the standard methodology.[14] and were mostly controlled for important confounders, age and gender, or age in male only or female only cohorts. Other potential confounders would include education, socioeconomic status, smoking and alcohol consumption. In the subgroup analysis, there was no significant difference in mortality risk between studies adjusting for age and gender or age only and studies additionally adjusting for such confounders (8 cohorts: pooled HR=1.04, 95%CI=1.03-1.05, p<0.001, I²=89%, 4 cohorts: pooled HR=1.04, 95%CI=1.03-1.04, p<0.001, I²=74%, respectively. P for subgroup difference=0.53). Lastly this is the first systematic review and meta-analysis focusing on FI and mortality.

There are several features of the FI which distinguish it from frailty phenotype. As mentioned above, the FI can evaluate frailty status in a graded manner, rather than just three frailty categorisations by frailty phenotype (robust, pre-frail and frail), and make a more precise risk prediction. Furthermore, those who have a missing value for specific frailty components may be excluded from analyses in frailty phenotype. However the FI can still be calculated by excluding missing deficits from both numerator and denominator, which is because deficits are considered to be interchangeable if a sufficiently large number of deficits are included.[42] Although one may argue that it is not practical in clinical settings to collect information of 30 or more health deficits to calculate the FI, most of the clinical information could be extracted from electronic medical record systems. A recent study created an electronic FI from readily available data in primary care electronic records and demonstrated robust predictive ability for mortality, hospitalisation and nursing home placement.[43]

This systematic review and meta-analysis was the first to quantitatively demonstrate the pooled mortality risk estimate according to frailty defined by the FI. Frailty measured by the FI is a strong predictor of death among older people as well as younger and middle-aged populations. A shorter follow-up period and lower female proportion seem to be associated with higher mortality risks according to frailty.

ABBREVIATIONS
CI: Confidence interval; CSHA: Canadian Study of Health and Aging; FI: Frailty Index; HR: Hazard ratio; OR: Odds ratio; SHARE: Survey of Health, Ageing and Retirement in Europe.

CONFLICT OF INTEREST
None.
REFERENCES
22. Mitnitski A, Song X, Rockwood K. Trajectories of changes over twelve years in the
Figure 1. Flow chart of systematic literature review

2611 studies identified through database searching
   Embase (n=1790)
   MEDLINE (n=561)
   CINAHL Plus (n=181)
   PsycINFO (n=79)

6 additional studies identified through other sources

Total of 2617 studies identified

651 duplicated studies excluded

1966 studies screened for titles and abstracts

1891 studies excluded by title and abstract screening

75 articles for full-text review

56 studies excluded by full-text review
   No HR/OR for mortality provided (n=25)
   Per groups based on FI (n=17)
   Per each additional deficit (n=4)
   Non-standard FI used (n=3)
   Same cohort used (n=3)
   Selected population (n=1)
   FI scale unknown (n=3)

19 studies for methodological quality assessment

19 studies for meta-analysis
Figure 2. Forest plots of mortality risk according to frailty measured by the Frailty Index.

A: Risk of dying (Hazard Ratio) per 0.01 increase in the Frailty Index score

B: Risk of dying (Hazard Ratio) per 0.1 increase in the Frailty Index score.

CI: Confidence interval, IV: inverse variance, NSHS: Nova Scotia Health Survey.

BWHHS: British Women’s Heart and Health Study, CI: Confidence interval, IV: inverse variance, MRC: MRC assessment study.
Table 1. Summary of included studies on Frailty Index and mortality.

<table>
<thead>
<tr>
<th>Author/Study</th>
<th>Year</th>
<th>Location</th>
<th>Sample size</th>
<th>Female (%)</th>
<th>Age (range)</th>
<th>Number of deficits</th>
<th>Follow-up period</th>
<th>Risk estimate HR/OR (95%CI)</th>
<th>Adjustment</th>
</tr>
</thead>
<tbody>
<tr>
<td>HR per 0.01 of FI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Searle Yale-PEP</td>
<td>2008</td>
<td>USA</td>
<td>754</td>
<td>64.6%</td>
<td>(72-98)</td>
<td>40</td>
<td>9 years</td>
<td>aHR=1.03 (1.02-1.04)</td>
<td>age, gender</td>
</tr>
<tr>
<td>Kulminski Cardiovascular Health Study</td>
<td>2008</td>
<td>USA</td>
<td>1,073</td>
<td>-</td>
<td>(≥65)</td>
<td>48</td>
<td>4 years</td>
<td>HR=1.049 (1.040-1.057)</td>
<td>unadjusted</td>
</tr>
<tr>
<td>Rockwood National Population Health Survey</td>
<td>2011</td>
<td>Canada</td>
<td>14,127</td>
<td>54.2%</td>
<td>(≥15)</td>
<td>42</td>
<td>14 years</td>
<td>aHR=1.04 (1.03-1.04)</td>
<td>age, gender, education</td>
</tr>
<tr>
<td>Yu Beijing Longitudinal Study of Aging (Urban sample)</td>
<td>2012</td>
<td>China</td>
<td>2,136</td>
<td>51.1%</td>
<td>(55-97)</td>
<td>35</td>
<td>8 years</td>
<td>aHR=1.042 (1.036-1.049)</td>
<td>age, gender, education</td>
</tr>
<tr>
<td>Yu Beijing Longitudinal Study of Aging (Rural sample)</td>
<td>2012</td>
<td>China</td>
<td>1,121</td>
<td>51.0%</td>
<td>(55-97)</td>
<td>35</td>
<td>8 years</td>
<td>aHR=1.041 (0.034-1.049)</td>
<td>age, gender, education</td>
</tr>
<tr>
<td>Bennett Chinese Longitudinal Healthy Longevity Survey</td>
<td>2013</td>
<td>China</td>
<td>6,300</td>
<td>53.0%</td>
<td>(≥50)</td>
<td>38</td>
<td>4 years</td>
<td>aHR=1.03 (1.03-1.04)</td>
<td>age, gender</td>
</tr>
<tr>
<td>Theou SHARE</td>
<td>2013</td>
<td>Europe*</td>
<td>36,306</td>
<td>54.6%</td>
<td>(≥80)</td>
<td>70</td>
<td>2 years</td>
<td>aHR=1.05 (1.05-1.06)</td>
<td>age, gender</td>
</tr>
<tr>
<td>Pena Nova Scotia Health Survey</td>
<td>2014</td>
<td>Canada</td>
<td>3,227</td>
<td>50.1%</td>
<td>(≥18)</td>
<td>23</td>
<td>10 years</td>
<td>aHR=1.04 (1.03-1.05)</td>
<td>age, gender</td>
</tr>
<tr>
<td>Blodgett EMAS</td>
<td>2016</td>
<td>Europe†</td>
<td>2,933</td>
<td>0%</td>
<td>(40-79)</td>
<td>39</td>
<td>4.4 years</td>
<td>aHR=1.07 (1.06-1.09)</td>
<td>age</td>
</tr>
<tr>
<td>Hao Project of Longevity and Aging in Duijiangyan</td>
<td>2016</td>
<td>China</td>
<td>767</td>
<td>68.0%</td>
<td>(90-108)</td>
<td>35</td>
<td>4 years</td>
<td>aHR=1.03 (1.02-1.04)</td>
<td>age, gender, education</td>
</tr>
<tr>
<td>Hoogendijk Longitudinal Aging Study Amsterdam</td>
<td>2016</td>
<td>Netherlands</td>
<td>2,218</td>
<td>-</td>
<td>(57-88)</td>
<td>32</td>
<td>19 years</td>
<td>aHR=1.03 (1.03-1.04)</td>
<td>age, gender</td>
</tr>
<tr>
<td>Author/Study</td>
<td>Year</td>
<td>Location</td>
<td>Sample size</td>
<td>Female (%)</td>
<td>Age (range)</td>
<td>Number of deficits</td>
<td>Follow-up period</td>
<td>Risk estimate HR/OR (95%CI)</td>
<td>Adjustment</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>------</td>
<td>----------</td>
<td>-------------</td>
<td>------------</td>
<td>-------------</td>
<td>-------------------</td>
<td>-----------------</td>
<td>----------------------------</td>
<td>------------------------------------------------</td>
</tr>
<tr>
<td>Miller NHANES</td>
<td>2016</td>
<td>USA</td>
<td>8,911</td>
<td>-</td>
<td>(20-)</td>
<td>46</td>
<td>8 years</td>
<td>aHR=1.03 (1.02-1.04)</td>
<td>age, gender</td>
</tr>
<tr>
<td>Mfinitski CSHA</td>
<td>2016</td>
<td>Canada</td>
<td>1,013</td>
<td>61.6%</td>
<td>80.8-65</td>
<td>61</td>
<td>6 years</td>
<td>aHR=1.041 (1.030-1.052)</td>
<td>age, gender</td>
</tr>
<tr>
<td>Kamaruzzaman BWHHS</td>
<td>2010</td>
<td>UK</td>
<td>4,286</td>
<td>100%</td>
<td>-</td>
<td>(60-79)</td>
<td>44</td>
<td>8.2 year</td>
<td>aHR=1.3 (1.2-1.4)</td>
</tr>
<tr>
<td>Kamaruzzaman MRC assessment study</td>
<td>2010</td>
<td>UK</td>
<td>11,195</td>
<td>59.9%</td>
<td>-</td>
<td>(≥75)</td>
<td>44</td>
<td>7.9 year</td>
<td>aHR=1.3 (1.2-1.3)</td>
</tr>
<tr>
<td>Theou CSHA</td>
<td>2012</td>
<td>Canada</td>
<td>2,305</td>
<td>62.1%</td>
<td>84.6 (70-105)</td>
<td>70</td>
<td>5 years</td>
<td>aHR=1.25 (1.20-1.30)</td>
<td>age, gender</td>
</tr>
<tr>
<td>Yu Beijing Longitudinal Study of Aging (Urban sample)</td>
<td>2012</td>
<td>China</td>
<td>2,136</td>
<td>51.1%</td>
<td>70.1 (55-97)</td>
<td>35</td>
<td>8 years</td>
<td>aHR=1.28 (1.23-1.32)</td>
<td>age, gender, education</td>
</tr>
<tr>
<td>Yu Beijing Longitudinal Study of Aging (Rural sample)</td>
<td>2012</td>
<td>China</td>
<td>1,121</td>
<td>51.0%</td>
<td>70.2-70.3 (55-97)</td>
<td>35</td>
<td>8 years</td>
<td>aHR=1.27 (1.21-1.32)</td>
<td>age, gender, education</td>
</tr>
<tr>
<td>Rivindrarajah EMAS</td>
<td>2013</td>
<td>Europe†</td>
<td>2,929</td>
<td>0%</td>
<td>59.9 (40-79)</td>
<td>39</td>
<td>4.3 years</td>
<td>aHR=1.49 (1.33-1.67)</td>
<td>age, center, smoking, partner status</td>
</tr>
<tr>
<td>Li GLOW</td>
<td>2014</td>
<td>Canada</td>
<td>3,985</td>
<td>100%</td>
<td>69.4 (≥55)</td>
<td>34</td>
<td>3 years</td>
<td>aOR=1.05 (1.03-1.06)</td>
<td>age, BMI, smoking, alcohol, education</td>
</tr>
<tr>
<td>Theou TILDA</td>
<td>2015</td>
<td>UK</td>
<td>4,961</td>
<td>54.2%</td>
<td>61.9 (≥50)</td>
<td>66</td>
<td>2 years</td>
<td>aOR=1.072 (1.040-1.106)</td>
<td>age, gender</td>
</tr>
<tr>
<td>Armstrong HAAS</td>
<td>2015</td>
<td>USA</td>
<td>3,845</td>
<td>0%</td>
<td>77.9 (72-91)</td>
<td>48</td>
<td>6 years</td>
<td>aOR=1.73 (1.57-1.92)</td>
<td>age, education</td>
</tr>
<tr>
<td>Li GLOW</td>
<td>2016</td>
<td>Canada</td>
<td>3,985</td>
<td>100%</td>
<td>69.4 (≥55)</td>
<td>34</td>
<td>3 years</td>
<td>aOR=1.33 (0.87-2.03)</td>
<td>age</td>
</tr>
</tbody>
</table>

* 15 European countries: Austria, Belgium, Czech Republic, Denmark, France, Germany, Greece, Ireland, Israel, Italy, Netherlands, Poland, Spain, Sweden, Switzerland
† 8 European countries: Belgium, Estonia, Hungary, Italy, Poland, Spain, Sweden, UK
95%CI= 95% confidence interval
(a)HR: (adjusted) Hazard ratio
(a)OR: (adjusted) Odd ratio
BMI: Body mass index
BWHHS: British Women’s Heart and Health Study
CSHA: Canadian Study of Health and Aging
EMAS: European Male Ageing Study
GLOW: Global Longitudinal Study of Osteoporosis in Women
FI: frailty index
NHANES: National Health and Nutrition Examination Survey
SES: Socioeconomic status
SHARE: Survey of Health, Ageing and Retirement in Europe
TILDA: The Irish Longitudinal study on Ageing
Yale-PEP: Yale Precipitating Events Project
Appendix 1.
Methodological quality assessment
Each of the eligible studies was further examined for methodological quality using the Newcastle-Ottawa scale for cohort studies.[21] This scale consists of nine items regarding selection (4 items), compatibility (2 items) and outcome (3 items) domains of cohort studies. The third item in the selection domain (ascertainment of exposure) was modified to confirm whether a study constructed the FI in accordance with the standardised method published by Searle et al.[14] The fourth one (demonstration that outcome of interest was not present at start of study) was not used in this study since the outcome of interest was mortality. A study was considered to have adequate quality of methodology and was included in the meta-analysis if four or more items out of eight were met by the modified scale.

Statistical analysis
The HR or OR along with 95% CI per 0.10 or 0.01 increase in the FI were extracted from the included studies and were used for the meta-analysis. The meta-analysis was conducted using the generic inverse variance method. Heterogeneity across the studies was assessed using Cochran’s Q statistic and I² statistic. When p value of Cochran’s Q statistic was less than 0.05, the studies were combined using a random-effects model. Otherwise a fixed-effects model was used. The studies with I² value of 25%, 50% and 75% were considered to have low, moderate and high degree of heterogeneity.[22] When significant heterogeneity was observed in the studies, its potential cause was explored by subgroup analysis and meta-regression analysis. Publication bias was assessed using Begg-Mazumdar’s and Egger’s tests and visually inspecting a funnel plot.

All statistical analyses were conducted using Review Manager 5 (version 5.2, The Cochrane Collaboration, Copenhagen, Denmark) and Comprehensive Meta-Analysis (version 3.3, Biostat, New Jersey, USA). The level of statistical significance was set at P<0.05.

Meta-regression and subgroup analysis
A high degree of heterogeneity was observed among 13 cohorts with HR of mortality per 0.01 increase in the FI and was explored using meta-regression analysis. Several study characteristics examined included publication year, location (Europe vs no Europe, US vs no US, Canada vs no Canada), sample size, female proportion, mean age, the number of deficits used for the FI, follow-up period, additional adjustment other than only age and gender and methodological quality score based on the modified eight-item Newcastle-Ottawa scale. Three[18, 34, 39] and Four[14, 18, 34, 39] studies did not report female proportion and mean age, respectively, and were not included in the analyses for each characteristic. The results suggested that two factors were significantly associated with higher mortality risks by the FI: (1) shorter follow-up periods (coefficient=-0.001, p=0.04, R² analog=0.24); and (2) lower female proportion of the studies (coefficient=-0.0005, p=0.00, R² analog=0.31). Appendix 5 A and B show the bubble plots for the follow-up periods and female proportion.

Heterogeneity of four cohorts with the follow-up periods of nine years or more decreased (I²=14%), while the high heterogeneity remained among nine studies with follow-up periods of eight years or less (I²=86%). Mortality risk according to frailty of the studies with follow-up of nine years or more was significantly lower than that of the studies with follow-up of eight years or less (p for difference=0.007). Excluding one male-only cohort[37] made little change to the high heterogeneity among the remaining 12 cohorts with mixed-gender populations (pooled HR=1.04, 95%CI=1.03-1.04, p<0.001, I²=83%).
*Publication Bias Assessment*

The 13 cohorts providing HR per 0.01 increase in the FI and six cohorts providing HR per 0.1 increase in the FI were assessed for publication bias. No significant publication bias was observed by Begg-Mazumdar’s (p=0.57 and 0.34, respectively) or Egger’s test (p=0.37 and 0.08, respectively). The funnel plots did not show obvious asymmetry. Begg-Mazumdar’s and Egger’s tests could not be done due to the small number of the included studies for the cohorts with OR per 0.01 increase of the FI (n=2) and the cohorts with OR per 0.1 increase of the FI (n=2).
## Appendix 2. A summary of the excluded studies examining mortality risk by the Frailty Index.

<table>
<thead>
<tr>
<th>Author/Study(Location)/Year</th>
<th>Sample size</th>
<th>Female (%)</th>
<th>Age (range)</th>
<th>Follow-up period</th>
<th>Number of deficits</th>
<th>How FI was used as a predictor variable</th>
<th>Effect measure for mortality risk (95%CI, p value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shi[1] BLSA (China) 2011</td>
<td>3,257</td>
<td>51.1%</td>
<td>(&gt;55)</td>
<td>8 years</td>
<td>35</td>
<td>per deficit</td>
<td>HR=1.13 (1.09-1.47) adjusted for age and gender.</td>
</tr>
<tr>
<td>Drubbel[2] (Netherlands) 2013</td>
<td>1,679</td>
<td>58.8%</td>
<td>73 (65-81)</td>
<td>2 years</td>
<td>36</td>
<td>per deficit</td>
<td>HR=1.166 (1.129-1.210, p=0.05) for combined outcomes (mortality, emergency department or out-of-hours GP surgery visits and nursing home admission), adjusted for age, gender and consultation gap.</td>
</tr>
<tr>
<td>Song[3] CSHA (Canada) 2014</td>
<td>7,239</td>
<td>59.9%</td>
<td>- (&gt;65)</td>
<td>10 years</td>
<td>42</td>
<td>per deficit</td>
<td>Age-adjusted OR=1.22 (1.18-1.26) for men and 1.14 (1.11-1.16) for women.</td>
</tr>
<tr>
<td>Yang[4] Chinese Longitudinal Healthy Longevity Survey (China) 2016</td>
<td>13,731</td>
<td>57.3%</td>
<td>- (&gt;65)</td>
<td>3 years</td>
<td>39</td>
<td>per deficit</td>
<td>HR by Weibull hazard models=1.04 to 1.10 in all age groups of 65-79, 80-89, 90-99, and 100+ (all p&lt;0.001) both in men and women, adjusted for age, ethnicity, residence, marital status, education, occupation, economic independence, economic status, co-residence with family, smoking, and exercise.</td>
</tr>
<tr>
<td>Bartley[5] Mayo Clinic Study of Aging (USA) 2016</td>
<td>2,356</td>
<td>49.8%</td>
<td>78.8 (70-89)</td>
<td>6.5 years</td>
<td>36</td>
<td>(i) per deficit (ii) 4 groups (cut-points: 0.10, 0.20, 0.30)</td>
<td>(i) HR=1.12 (1.10-1.15, p&lt;0.001) adjusted for age, gender and education. (ii) HR=1.47 (1.03-2.10, p=0.03), 2.65 (1.86-3.78, p&lt;0.001) and 3.91 (2.69-5.68, p&lt;0.001) for groups 0.11-0.20, 0.21-0.30 and &gt;0.30, respectively (reference group= 0-0.10).</td>
</tr>
<tr>
<td>Hyde[6] Australia (Aboriginal Australians) 2016</td>
<td>363</td>
<td>54.5%</td>
<td>60.7 (45-96)</td>
<td>6.7 years</td>
<td>20</td>
<td>(i) per deficit (ii) 2 groups (cut-point 0.2)</td>
<td>(i) HR=1.14 (1.1-1.2) adjusted for age and gender. (ii) HR=1.9 (1.2-3.0) adjusted for age and gender.</td>
</tr>
<tr>
<td>Lucicesare[7] Conseilie Study of Brain Aging (Italy) 2010</td>
<td>1,016</td>
<td>55.4%</td>
<td>74.7 (&gt;65)</td>
<td>4 years</td>
<td>43</td>
<td>apparently 2 groups (cut-point 0.25)</td>
<td>HR=5.26 (1.05-26.42, p=0.04) adjusted for age, gender and Conseilie Study of Brain Aging score.</td>
</tr>
<tr>
<td>Tang[8] BLSA (China) 2013</td>
<td>3,257</td>
<td>51.1%</td>
<td>70.1 (&gt;55)</td>
<td>15 years</td>
<td>35</td>
<td>2 groups (cut point: 0.22)</td>
<td>HR=2.06 (1.82-2.32, p&lt;0.01) adjusted for age, gender and education.</td>
</tr>
<tr>
<td>Widagdo[9] Australian Longitudinal Study of Ageing (Australia) 2015</td>
<td>2,087</td>
<td>49.4%</td>
<td>78.2 (&gt;65)</td>
<td>3 years</td>
<td>39</td>
<td>2 groups (frailty or not) (cut-point 0.25)</td>
<td>OR=3.2 (2.4-4.1).</td>
</tr>
<tr>
<td>Kulminski[10] Cardiovascular Health Study (USA) 2008</td>
<td>4,721</td>
<td>-</td>
<td>- (&gt;65)</td>
<td>4 years</td>
<td>48</td>
<td>3 groups (robust, prefrail, frail) (cut-points: 0, 0.4)</td>
<td>Unadjusted HR=1.94 (1.45-2.61) for prefrail and 4.45 (3.26-6.08) for frail. (reference group: robust).</td>
</tr>
<tr>
<td>Author/Study(Location)/Year</td>
<td>Sample size</td>
<td>Female (%)</td>
<td>Age (range)</td>
<td>Follow-up period</td>
<td>Number of deficits</td>
<td>How FI was used as a predictor variable</td>
<td>Effect measure for mortality risk (95%CI, p value)</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-------------</td>
<td>------------</td>
<td>-------------</td>
<td>------------------</td>
<td>------------------</td>
<td>----------------------------------------</td>
<td>-------------------------------------------------</td>
</tr>
<tr>
<td>Malmstrom[11] (USA) 2014</td>
<td>998</td>
<td>-</td>
<td>(49-65)</td>
<td>9 years</td>
<td>25</td>
<td>3 groups (robust, prefrail, frail) (cut-points: 0.20, 0.25)</td>
<td>OR=1.77 (0.92-3.41, p=0.08) for prefrail and 2.28 (1.46-3.55, p&lt;0.001) for frail adjusted for age and gender (reference group=robust)</td>
</tr>
<tr>
<td>Song[12] National Population Health Survey of Canada (Canada) 2010</td>
<td>2,740</td>
<td>60.8%</td>
<td>74.0 (65-102)</td>
<td>10 years</td>
<td>36</td>
<td>3 groups (cut-points: 0.08, 0.25)</td>
<td>HR=1.57 (1.41-1.74) adjusted for age and gender.</td>
</tr>
<tr>
<td>Wang[13] BLSA (China) 2013</td>
<td>3,257</td>
<td>51.1%</td>
<td>(≥55)</td>
<td>15 years</td>
<td>28</td>
<td>3 groups (cut-points: 0.08, 0.15)</td>
<td>Higher frailty levels associated with higher mortality risk in both smokers and non-smokers.</td>
</tr>
<tr>
<td>Li[14] Global longitudinal study of osteoporosis in women (Canada) 2015</td>
<td>3,985</td>
<td>100%</td>
<td>69.4 (≥55)</td>
<td>3.01 years</td>
<td>34</td>
<td>(i) 3 groups (cut-points: 0.20, 0.35) (ii) 3 groups (mean: 0.18, 0.29, 0.35) (iii) 5 groups (cut-points: 0.14, 0.28, 0.42, 0.56))</td>
<td>(i) HR=1.95 (1.06-3.61) for intermediate frailty and 4.26 (2.34-7.76) for high frailty. (ii) HR=2.46 (1.39-4.36) for intermediate frailty and 478 (2.65-8.63) for high frailty. (iii) HR=1.81 (1.46-2.24) with each increment in FI grouping. All models adjusted for age, smoking, alcohol, BMI and education.</td>
</tr>
<tr>
<td>Clegg[15] Health Improvement Network databases (UK) 2016</td>
<td>207,720</td>
<td>55%</td>
<td>(65-95)</td>
<td>1, 3, 5 years</td>
<td>36</td>
<td>4 groups (cut-points: 0.12, 0.24, 0.36)</td>
<td>1, 3 and 5 year-mortality HR=1.66-1.92, 2.54-3.10 and 3.83-4.52 adjusted for age and gender for groups &gt;0.12-0.24, &gt;0.24-0.36 and &gt;0.36, respectively (reference group=0-0.12).</td>
</tr>
<tr>
<td>Gu[16] Chinese Longitudinal Healthy Longevity Survey (China) 2009</td>
<td>13,861</td>
<td>57.2%</td>
<td>(65-109)</td>
<td>3 years</td>
<td>39</td>
<td>4 groups (quartile)</td>
<td>HR by Weibull hazard models=1.18-2.12 for 2nd quartile, 1.55-2.38 for 3rd quartile and 2.41-4.56 for 4th quartile, stratified by age and gender adjusted for age ethnicity, residence, socioeconomic status, family/social connection and support and health practices (reference group=1st quartile).</td>
</tr>
<tr>
<td>Fang[17] BLSA (China) 2012</td>
<td>3,257</td>
<td>51.1%</td>
<td>70.1 (≥55)</td>
<td>8 years</td>
<td>33</td>
<td>5 groups (cut points: 0.03, 0.10, 0.20, 0.50)</td>
<td>OR=1.50 (1.41-1.60) adjusted for age, gender and education. HR=1.29 (1.25-1.33) adjusted for age, gender, education, falls and fractures.</td>
</tr>
<tr>
<td>Garcia-Gonzalez[18] Mexican Health and Aging Study (Mexico) 2009</td>
<td>4,082</td>
<td>52.5%</td>
<td>73 (≥65)</td>
<td>2 years</td>
<td>34</td>
<td>5 groups (cut-points: 0.07, 0.14, 0.21, 0.35)</td>
<td>HR=0.93 (0.58-1.50), 1.56 (1.00-2.44), 2.20 (1.42-3.41), 6.45 (4.10-10.14) for 2nd, 3rd, 4th and 5th groups adjusted for age and gender (reference group=1st group).</td>
</tr>
<tr>
<td>Saum[19] ESTHER (Germany) 2014</td>
<td>9,886</td>
<td>54.9%</td>
<td>62.0 (50-75)</td>
<td>8.7 years</td>
<td>34</td>
<td>5 groups (tertiles)</td>
<td>HR=1.08 (0.84-1.39), 1.32 (1.05-1.66), 1.77 (1.41-2.22) and 2.60 (2.11-3.20) for 2nd, 3rd, 4th and 5th quintile adjusted for age, gender and smoking (reference group=1st quintile).</td>
</tr>
<tr>
<td>Armstrong[20] Honolulu-Asia Aging Study (USA) 2015</td>
<td>3,801</td>
<td>0%</td>
<td>77.9 (71-93)</td>
<td>21 years</td>
<td>36</td>
<td>6 groups (cut-points: 0.05, 0.15, 0.25, 0.35, 0.5)</td>
<td>HR=1.44 (1.39-1.49) with each increment in FI grouping.</td>
</tr>
<tr>
<td>Author/Study(Location)/Year</td>
<td>Sample size</td>
<td>Female (%)</td>
<td>Age (range)</td>
<td>Follow-up period</td>
<td>Number of deficits</td>
<td>How FI was used as a predictor variable</td>
<td>Effect measure for mortality risk (95%CI, p value)</td>
</tr>
<tr>
<td>----------------------------</td>
<td>-------------</td>
<td>------------</td>
<td>-------------</td>
<td>------------------</td>
<td>-------------------</td>
<td>----------------------------------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Jones[21] CSHA (Canada) 2005</td>
<td>3,736</td>
<td>38.3%</td>
<td>(≥65)</td>
<td>5 years</td>
<td>14</td>
<td>7 groups (cut-points: 0.23, 0.31, 0.40, 0.48, 0.60, 0.74)</td>
<td>HR=1.23 (1.18-1.29) with each increment in FI grouping adjusted for age, gender and education.</td>
</tr>
<tr>
<td>Mitnitski[22] CSHA (Canada) 2011</td>
<td>2,305</td>
<td>62.1%</td>
<td>(&gt;70)</td>
<td>5 years</td>
<td>47</td>
<td>7 groups (not specified)</td>
<td>OR=1.56 adjusted for age, gender and baseline cognitive error state.</td>
</tr>
<tr>
<td>Howlett[23] CSHA (Canada) 2014</td>
<td>1,013</td>
<td>-</td>
<td>(≥65)</td>
<td>6 years</td>
<td>61</td>
<td>per 0.01 increase of FI</td>
<td>HR=1.04 (1.03-1.05) per 0.01 increase adjusted for age and gender.</td>
</tr>
<tr>
<td>Davis[24] CSHA (Canada) 2011</td>
<td>1,295</td>
<td>-</td>
<td>(≥65)</td>
<td>5 years</td>
<td>not shown</td>
<td>per 0.01 increase of FI</td>
<td>HR=1.04 (1.02, 1.06, p&lt;0.05) adjusted for age and gender.</td>
</tr>
<tr>
<td>Gu[25] CLHLS (China) 2015 (Female)</td>
<td>3,557</td>
<td>100%</td>
<td>(&gt;100)</td>
<td>3.7 years</td>
<td>39</td>
<td>per 0.01 increase of FI</td>
<td>HR=1.016 (1.014-1.018) adjusted for “demographics, socioeconomic status, and health practice”</td>
</tr>
<tr>
<td>Gu[25] CLHLS (China) 2015 (Male)</td>
<td>877</td>
<td>0%</td>
<td>(&gt;100)</td>
<td>3.7 years</td>
<td>39</td>
<td>per 0.01 increase of FI</td>
<td>HR=1.014 (1.010-1.018) adjusted for “demographics, socioeconomic status, and health practice”</td>
</tr>
<tr>
<td>Song[26] CSHA (Canada) 2007</td>
<td>8,547</td>
<td>59.5%</td>
<td>(≥65)</td>
<td>6 years</td>
<td>40</td>
<td>“each increment in the FI”</td>
<td>HR=1.38 (1.14-1.72) and 1.18 (1.11-1.26) in rural and urban participants, respectively.</td>
</tr>
<tr>
<td>Kulminski[27] Framingham Heart Study (USA) 2008</td>
<td>5,882</td>
<td>59.7%</td>
<td>(44-88)</td>
<td>24 years</td>
<td>39</td>
<td>not shown</td>
<td>HR=1.62 (1.53-1.71) adjusted for age, gender, smoking and BMI.</td>
</tr>
<tr>
<td>Rockwood[28] CSHA (Canada) 2005</td>
<td>2,305</td>
<td>-</td>
<td>(≥65)</td>
<td>5 years</td>
<td>70</td>
<td>not shown</td>
<td>HR=1.26 (1.24-1.29) adjusted for age, gender and education.</td>
</tr>
</tbody>
</table>

BLSA: Beijing Longitudinal Study of Ageing
CI: Confidence interval
CLHLS: Chinese Longitudinal Health and Longevity Study
CSHA: Canadian Study of Health and Aging
FI: Frailty index
HR: Hazard ratio
OR: Odds ratio
Reference

20. Armstrong JJ, Mitnitski A, Launer LJ, White LR, Rockwood K. Frailty in the Honolulu-Asia Aging Study: deficit accumulation in a male cohort followed to 90%
Appendix 3. References of the included studies.
Appendix 4. Forest plots of mortality risk according to frailty measured by the Frailty Index.

A: Odds of dying (Odds Ratio) per 0.01 increase in the Frailty Index score

B: Odds of dying (Odds Ratio) per 0.1 increase in the Frailty Index score.

CI: Confidence interval, IV: inverse variance
Appendix 5. Bubble plots for the follow-up periods (A) and female proportion (B)

A

Regression of Log hazard ratio on Follow-up period

B

Regression of Log hazard ratio on Female proportion