The frailty of constructs and the construct of frailty in geriatric medicine and research

Quoc Dinh Nguyen
Department of Epidemiology, Biostatistics and Occupational Health,
McGill University, Montreal

October 2021

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy in epidemiology

Table of contents

Abstract

Frailty has been defined as a “state of increased vulnerability” which increases the risk of adverse outcomes. Although there is currently no consensus definition, frailty has consistently been associated with death, institutionalization, and disability. These associations have driven strong, but mostly unheeded, calls from the research and clinical community to include frailty in clinical care processes. Frailty is one of many nosological constructs that have been developed in response to the complexity of aging. Geriatric research and practice rely on age-related constructs which diverge from the traditional biomedical model of disease: frailty, multimorbidity, and other geriatric syndromes are used to describe and characterize the variability in the health status of older adults. Despite vigorous interest for such constructs, issues related to their definition, reliability, and applicability in the context of clinical practice have not been well explored.

The overarching objective for this dissertation was to explore the conceptual underpinning and clinical applicability of age-related constructs, with an emphasis on frailty. The first objective was to examine whether older adults show increasing heterogeneity as they age, as a motivation for developing age-specific constructs. A novel application of methods was used to disentangle the within-age and between-age variability of 34 health characteristics within eight domains in 30,097 participants from the Canadian Longitudinal Study on Aging. I demonstrated that heterogeneity increases with aging, but not for all health characteristics and domains, and not uniformly. Clinical implications and research opportunities of heterogeneity include the importance of the comprehensive geriatric assessment, of measurement and scaling of variables, and the need to develop new multidimensional constructs that distinguish older adults among themselves.

The second objective was to examine the measurement of age-related constructs. I investigated the reliability and clinical correlates of reliability of the deficit-accumulation frailty index. Monte Carlo methods were used to simulate 12,000 studies comparing various implementations of the frailty index in 12,080 participants 65 years and older from the Canadian Longitudinal Study on Aging. I showed that the number and composition of items of individual FIs strongly influence their reliability. Descriptive estimates using frailty indices and predictive estimates between frailty indices and mortality varied markedly between implementations. This lack of reliability and stability of estimates lowers the generalizability and clinical application of study findings based on frailty indices.

The third objective focused on determining the applicability of study results to clinical practice. A literature search was conducted to identify 26 existing frameworks appraising study "applicability." I analyzed and synthesized frameworks and criteria according to the scope and level of aggregation of the evidence appraised, the target user, and the specific area of applicability. Since available frameworks did not distinguish study results by *clinical* applicability, a novel framework to appraise clinical applicability was proposed which categorizes studies into three evidence domains (research domain, practice informing, and practice changing) using six criteria (Validity, Indication-informativeness, Clinical relevance, Originality, Risk-benefit comprehensiveness, and Transposability, VICORT).

This framework was used for the fourth objective to appraise the clinical applicability of recently published articles on frailty. A mapping review was conducted by systematically sampling 476 articles published in 2017–2018 and investigating whether these articles informed practice, changed practice, or belonged in the research domain. Among all articles, 63 (13%) articles were categorized as practice informing, 11 (2%) as potentially practice changing, and 1 (0.2%) as clearly

practice changing. The lack of indication-informativeness (96%) and originality (83%) were the most important reasons hampering clinical applicability. Recommendations are proposed for future research on frailty, which may also extend to other age-related constructs.

Résumé

La fragilité est définie comme un « état de vulnérabilité accru » qui augmente le risque d'issues défavorables. Bien qu'il n'y ait, à l'heure actuelle, aucune définition de la fragilité qui fasse consensus, la fragilité a été associée de façon soutenue à la mortalité, à l'institutionnalisation et aux incapacités. Ces associations ont poussé des chercheurs et des cliniciens à proposer l'inclusion du concept de la fragilité aux trajectoires et aux processus cliniques. La fragilité est l'un des nombreux concepts nosologiques développés en réponse à la complexité du vieillissement. La recherche en gériatrie et la pratique de la gériatrie reposent sur des concepts nosologiques liés à l'âge qui divergent du modèle biomédical traditionnel des maladies : la fragilité, la multimorbidité et d'autres syndromes gériatriques sont utilisés pour décrire et caractériser la variabilité des états de santé des personnes âgées. Malgré le grand intérêt porté à ces concepts, des enjeux liés à leur définition, leur fiabilité et leur applicabilité, en contexte de pratique clinique, n'ont pas été dûment explorés.

L'objectif général de cette thèse est d'explorer les fondements et l'applicabilité clinique des concepts nosologiques liés à l'âge, en particulier la fragilité. Le premier objectif est d'examiner si les personnes âgées démontrent une hétérogénéité grandissante alors qu'elles vieillissent, ce qui sous-tendrait le développement de concepts nosologiques spécifiques à l'âge. Une application innovante de méthodes statistiques a été employée pour distinguer les variabilités intra-âge et inter-âge de 34 caractéristiques de santé comprises dans huit domaines, au sein de 30 097 participants dans l'Étude longitudinale canadienne sur le vieillissement. Je démontre que l'hétérogénéité augmente avec l'âge chronologique, mais pas pour toutes les caractéristiques ni domaines, et de façon non uniforme. Les répercussions cliniques et en recherche de ces résultats incluent l'importance de l'évaluation gériatrique complète, des propriétés de la mesure des variables et la

nécessité de développer de nouveaux construits multidimensionnels qui différencient les personnes âgées entre elles.

Le deuxième objectif est d'examiner les propriétés de la mesure des concepts liés à l'âge. J'analyse la fiabilité et les corrélats de la fiabilité de l'index de fragilité basé sur l'accumulation des déficits. Des méthodes Monte Carlo ont été utilisées pour simuler 12 000 études comparant des configurations variées de l'index de fragilité au sein de 12 080 participants de 65 ans et plus dans l'Étude longitudinale canadienne sur le vieillissement. Je démontre que le nombre et la composition des items constituant les déficits de chaque index de fragilité influencent grandement leur fiabilité. Les estimés descriptifs utilisant les index de fragilité et les estimés prédictifs de la fragilité pour la mortalité varient selon les configurations. La faible fiabilité des estimés réduit la généralisabilité et l'applicabilité clinique des résultats d'études basées sur les index de fragilité.

Le troisième objectif vise à déterminer l'applicabilité à la pratique clinique des résultats d'études. Une recherche de la littérature a été réalisée pour identifier 26 cadres existants qui s'intéressent à l'applicabilité clinique. J'ai analysé et fait la synthèse des cadres et de leurs critères selon la portée et le niveau d'agrégation des données probantes, l'utilisateur ciblé et la thématique d'applicabilité. Comme les cadres existants ne différencient pas les résultats d'études selon leur applicabilité *clinique*, un nouveau cadre est proposé pour catégoriser les études en trois domaines de données probantes (domaine de la recherche, informant la pratique, changeant la pratique) en s'appuyant sur six critères (validité, informe une indication, pertinence clinique, originalité, analyse exhaustive des risques et des bénéfices et transposabilité, VICORT [acronyme en anglais]).

Ce cadre a été utilisé pour le quatrième objectif d'évaluer l'applicabilité clinique des études récentes portant sur la fragilité. Une revue exploratoire a été réalisée en échantillonnant de façon

systématique 476 articles publiés en 2017-2018 et en déterminant si ces articles informaient la pratique, changeaient la pratique ou appartenaient au domaine de la recherche. Parmi tous les articles, 63 (13%) ont été catégorisés comme informant la pratique, 11 (2%) comme changeant potentiellement la pratique et 1 (0.2%) comme changeant clairement la pratique. Les deux raisons principales entravant l'applicabilité clinique sont un manque d'information sur une indication clinique (96%) et d'originalité (83%). Des recommandations sont proposées afin d'améliorer la recherche future portant sur la fragilité et peuvent aussi s'appliquer à d'autres concepts nosologiques liés à l'âge.

Acknowledgements

Few lengthy but time-bound endeavors in life are guided by the same core of people. A PhD being one of them, there are a few people I wish to acknowledge who have contributed to the journey and the destination. Counterfactuals are always ill-defined, and though well-defined interventions may be hard to identify, they have undoubtedly altered the course of my doctoral training, production, and learnings.

My doctoral supervisors and committee member, Drs. Christina Wolfson, Erica Moodie, and Mark Keezer. Tina, what started out (at the Allen Memorial) as an appreciation for your open-mindedness became one for your signature mix of spiritedness, kindness, and wisdom. I have been lucky to be a beneficiary, and I hope to pay it forward in the coming years. Erica, with your laser-sharp focus, clarity of mind, and concern for students, you have taught me that it is often better to un-Proust myself and Hemingway: thank you. Mark, I am grateful for your pragmatic blend of good nature, humor, and experience as a clinician and researcher.

I wish to extend my recognition to my classmates and the Epidemiology, Biostatistics, and Occupational Health Faculty for sharing, generating, and dissecting ideas that have been foundational for my thesis. Specifically, Dr. Robert Goulden, for invigorating yet anchored discussions on epidemiology, medicine, and beyond; Dr. Sahar Saeed, for both indulging and shaking me on my grandiose pitches and allowing them to be as multilevel as they were.

To my parents, for their deep commitment to provide my sisters and me with every opportunity that they could summon and for their unwavering support for our education. Finally, to my wife Dr. Andréanne Wassef, for being my sounding board, my rock, and my colors for the pursuits of my (our) life, of which this thesis is merely one example.

Statement of financial support

I would like to acknowledge the financial support for my doctoral training from the Fonds de recherche du Québec-Santé, the Canadian Institutes of Health Research, the Fondation du Centre hospitalier de l'Université de Montréal, and the Department of Medicine of the Université de Montréal.

Contribution of authors

Manuscript 1. Quoc Dinh Nguyen, Erica M. Moodie, Marie-France Forget, Philippe Desmarais, Mark R. Keezer, Christina Wolfson. Health heterogeneity in older adults: exploration in the Canadian Longitudinal Study on Aging. *Journal of the American Geriatrics Society*. 2021;69(3):678–687.

Manuscript 2. Quoc Dinh Nguyen, Erica M. Moodie, Mark R. Keezer, Christina Wolfson. Clinical correlates and implications of the reliability of the frailty index in the Canadian Longitudinal Study on Aging. *Published electronically ahead of print in the Journals of Gerontology: Series A, Medical Sciences (https://doi.org/10.1093/gerona/glab161).*

Manuscript 3. Quoc Dinh Nguyen, Erica M. Moodie, Philippe Desmarais, Robert Goulden, Marie-France Forget, Eric Peters, Sahar Saeed, Mark R. Keezer, Christina Wolfson. Appraising clinical applicability of studies: mapping and synthesis of current frameworks, and proposal of the FrACAS framework and VICORT checklist. S*ubmitted to BMC Medical Research Methodology.*

Manuscript 4. Quoc Dinh Nguyen, Erica M. Moodie, Philippe Desmarais, Marie-France Forget, Han Ting Wang, Mark R. Keezer, Christina Wolfson. The state of frailty in research: a mapping review of its clinical applicability to practice. *Submitted to Ageing Research Reviews.*

As the first author of the four thesis manuscripts, I was responsible for conceiving the research questions, designing the studies, conducting statistical and qualitative analyses, and drafting the manuscripts. I received guidance from Christina Wolfson, Erica M. Moodie, and Mark R. Keezer for the design, analysis, and interpretation of data, drafting, and revision of all four manuscripts. Marie-France Forget and Philippe Desmarais assisted in the analysis and interpretation of data for

Manuscript 1. Marie-France Forget, Philippe Desmarais, Robert Goulden, Eric Peters, and Sahar Saeed contributed to the qualitative analysis and interpretation of data for Manuscript 3. Philippe Desmarais, Marie-France Forget, and Han Ting Wang contributed to the qualitative analysis and interpretation of data for Manuscript 4. All co-authors contributed significantly to the content, critically reviewed, and provided input during manuscript revision.

Statement of originality

The work presented in this dissertation constitutes original scholarship advancing knowledge of heterogeneity in aging in relation with frailty and age-related nosological constructs. Findings presented further the knowledge on the applicability of frailty to clinical practice and proposes avenues to improve the development of age-related nosological constructs. Manuscript 1 is the first wide-ranging exploration of heterogeneity in health characteristics in a contemporary population of adults and uses innovative statistical methods. It has been published in the *Journal of the American Geriatrics Society* accompanied by an editorial by Luigi Ferrucci and George Kuchel. Manuscript 2 is the first study using Monte Carlo methods to investigate the reliability of frailty indices and to produce clinically oriented descriptive and predictive estimates. It has been accepted for publication in the *Journals of Gerontology, Series A: Medical Sciences*. Manuscript 3 proposes a novel framework and criteria to appraise *clinical* applicability of studies and complements existing quality, general applicability, and generalizability appraisal frameworks. Manuscript 4 is the first mapping review on frailty which sought to deliberately examine clinical applicability and reasons that might hamper translation to practice.

While I received guidance from my supervisors, committee member and co-authors on the substantive, methodological, and statistical aspects of this thesis, I declare that the conception, execution, and drafting of the work in this dissertation are my own.

List of Tables

List of Figures

Acronyms and abbreviations

5MWT	5-minute walk test
ACS	Acute coronary syndrome
ADL	Activities of daily living
AGREE	Appraisal of guidelines, research and evaluation
ALT	Alanine aminotransferase
ASCVD	Atherosclerotic cardiovascular disease
ASTAIRE	Analyse de la transférabilité et accompagnement à l'adaptation des interventions en promotion de la santé
AVR	Aortic valve replacement
BMD	Bone mineral density
BMI	Body mass index
CES-D	Center for Epidemiologic Studies—Depression Scale
CFS	Clinical frailty scale
CGA	Comprehensive geriatric assessment
CHF	Chronic heart failure
CSHA	Canadian Study of Health and Aging
CLSA	Canadian Longitudinal Study on Aging
COPD	Chronic obstructive pulmonary disease
CPG	Clinical practice guideline
CRP	High sensitivity C-reactive protein
CVA	Cerebrovascular accident
EFS	Edmonton Frail Scale
EVAT	External validity assessment tool
FrACAS	Framework appraising the clinical applicability of studies
FEV1	Forced expiratory volume in 1 second
FI	Deficit-accumulation frailty index
FI-CGA	Frailty index – Comprehensive geriatric assessment
FI-LAB	Frailty index – Laboratory
FRAIL	Fatigue, resistance, ambulation, illness, loss of weight
GFR	Glomerular filtration rate
GI	Gastrointestinal
GRADE	Grading of recommendations, assessment, development and evaluations
GRADE EtD	GRADE evidence to decision
GRASP	Grading and assessment of predictive tools
HbA1C	Hemoglobin A1C
HDL	High-density lipoprotein
HFRS	Hospital frailty risk score
HRQoL	Health-related quality of life
HSR	Health services research
HTA	Health technology assessment
iADL	Instrumental activities of daily living
ICC_A	Intraclass correlation coefficient for agreement
ICU	Intensive care unit
IQR	Interquartile range

ISAR	Identification of seniors at risk
ISAR-HP	ISAR – hospitalized patients
ISAT	Intervention scalability assessment tool
LDL	Low-density lipoprotein
LMIC	Low- and middle-income country
LSA	Life Space Assessment
mFI-5	Modified frailty index – 5 items
mFI-11	Modified frailty index – 11 items
mmHg	Millimeters of mercury
MoCA	Montreal cognitive assessment
PABAK	Prevalence-adjusted bias-adjusted kappa
PASE	Physical activity scale for the elderly
PFP	Fried physical frailty phenotype
PRECIS	Pragmatic–explanatory continuum indicator summary
PRISMA-7	Programme de recherche sur l'intégration des services pour le maintien de l'autonomie
PROBAST	Prediction model risk of bias assessment tool
PR-Tool	PRECIS-Review tool
OARS	Older Americans resources and services
RCT	Randomized controlled trial
RE-AIM	Reach, effectiveness, adoption, implementation, maintenance
RoB	Risk of bias
RoBINS-I	Risk of bias in non-randomised studies - of interventions
SEm	Standard error of measurement
SHARE-FI	Survey of health, ageing and retirement in Europe frailty instrument
SOF	Study of osteoporotic fractures
STP	SUPPORT Tools for evidence-informed health policymaking
TAVI	Transcatheter aortic valve implantation
TAVR	Transcatheter aortic valve replacement
TFI	Tilburg Frailty Indicator
TIA	Transient ischemic attack
TSH	Thyroid-stimulating hormone
TUG	Timed up and go
VES-13	Vulnerable elders survey-13
VICORT	Validity, indication-informativeness, clinical relevance, originality, risk-benefit comprehensiveness, transposability
VMS	VeiligheidsManagementSysteem
WBC	White blood cell

Chapter 1. Introduction

1.1 Overview and significance

In the last fifty years, epidemiological research on aging and geriatric medicine have set their sights on frailty. Frailty has been defined as a "state of increased vulnerability" which increases the risk of adverse outcomes.[1] Research on frailty has blossomed in the last two decades, yet no consensus definition of frailty currently exists, and the multiple frailty definitions in use do not necessarily identify the same individuals as having frailty.[2–5] Even if measured using various instruments, frailty has been shown to be associated with adverse outcomes, such as death, institutionalization, and disability.[6–8] These consistent associations have driven strong, but mostly unheeded, calls from the research and clinical community to include the assessment of frailty in the clinical care process.[9–11] Beyond the lack of consensus definition, reasons for the relative disconnect between the interest in frailty in research and its translation to practice have not been well described.

Frailty is one of many age-related constructs concerned with the classification of diseases, i.e., *nosological* constructs, used in research on aging and geriatric medicine. Multimorbidity,[12,13] polypharmacy,[14] delirium[15,16] and other geriatric syndromes[17,18] are among constructs identified in response to the variability and complexity in health states of older adults. By design, frailty and these constructs do not directly fit the biomedical model of disease that assumes a single causal factor and underlies most of clinical epidemiology and medicine.[19] What has been gained or lost by diverging from the standard model to more flexibly account for the particularities of geriatric research and medicine remains unexplored. Epidemiology faces the dual challenge of accurately describing health states and determining their causal contribution of, and to, various health variables, ultimately to inform interventions that improve health outcomes in population and

individuals.[20] A tension may exist between the most accurate *description* of complex phenomena in older adults under the lens of epidemiology and their *usefulness and applicability* in clinical practice.

In this dissertation, I explored the continuum between the description and applicability of nosological constructs, at the intersection of epidemiology and clinical practice. Using frailty as an exemplar of an age-related construct,[21] I used quantitative methods to examine (i) the basis underlying age-related constructs and (ii) the ability of a frailty measure to reliably characterize individuals as having frailty. I further used qualitative methods to (iii) identify criteria by which studies featuring constructs can successfully inform clinical practice and (iv) appraise whether the construct of frailty has recently been successful in informing practice. Each manuscript examined a specific facet of translating constructs from description to clinical application. This dissertation is a probe into frailty, constructs, and the frailty of constructs that are used in geriatric epidemiology and medicine.

1.2 Research objectives

The overarching goal of this research was to explore the conceptual underpinning and clinical applicability of age-related nosological constructs, with an emphasis on frailty. There are four objectives, each of which is addressed in a manuscript which forms the basis of a chapter of this thesis:

1. To explore whether, and how, heterogeneity increases in individuals from middle to advanced age, as a basis for developing constructs related to aging (Manuscript 1).

2. To investigate the reliability of frailty indices and the stability of frailty-index derived estimates when computed in a single community-dwelling older adult population (Manuscript 2)
3. To identify criteria and to develop a tool to appraise, classify, and improve applicability of a research study to clinical practice (Manuscript 3).
4. To characterize the overall clinical applicability of recent frailty research and potential limits to applicability (Manuscript 4).

1.3 Organization of thesis

The format of this thesis is manuscript-based. Chapter 2 provides a literature review and background on frailty: the multiple definitions of frailty, a brief history of its development, an overview of potential frailty pathophysiological mechanisms, and a presentation of three dominant frailty frameworks. A contextualization of frailty and its purpose within the broader scope of aging, geriatrics, epidemiology, and medicine follows. Chapter 3 presents a summary of the data source and methods that will be applied in the thesis. Chapter 4 presents the exploration of heterogeneity in older adults, i.e., the underlying basis for age-related constructs (Manuscript 1). After examining heterogeneity in older adults as the basis for frailty and other constructs, Chapter 5 investigates whether a well-established and influential frailty measure can reliably identify individuals as having frailty (Manuscript 2). Chapter 6 examines the clinical function of nosological constructs: a literature search is conducted to identify criteria on which clinical applicability of studies can be determined; a novel framework is proposed (Manuscript 3). Chapter 7 takes elements from this proposed framework to formally appraise the clinical applicability of recently published frailty articles (Manuscript 4). Chapters 4 to 7 each have their own references. Implications and recommendations are discussed within each chapter. Chapter 8 summarizes and synthesizes

findings, strengths and limitations, and overarching implications for frailty. The thesis concludes with a brief discussion on the ontology of constructs and future directions for research.

Chapter 2. Background and scope

2.1 FRAILTY

2.1.1 The (con)current definition(s) of frailty

Frailty has been defined as "a state of increased vulnerability to poor resolution of homeostasis after a stressor event, which increases the risk of adverse outcomes, including falls, delirium, and disability."[1] Despite its longstanding prominence in research on aging, there is currently no standard or consensus definition of frailty.[2,3] A recent review on frailty instruments identified 67 different frailty instruments, of which nine were systematically characterized and included differing domains in their composition (i.e., physical function, disability, physical activity, cognition, comorbidity, weight loss, and others).[7] The lack of uniformity in the elements used to identify frailty was also reported in another systematic review of 22 articles: physical function, mobility, and cognition were assessed in half of more of instruments, but self-rated health, age, urinary incontinence, or use of health services were measured in less than a third.[22]

Underlying all existing definitions or operationalizations of frailty is the core idea of vulnerability, associated with multiple age-related processes from multiple domains and with adverse outcomes.[23,24] Amid debates on the nature and definition of frailty,[25] a relatively consistent conceptualization of frailty suggests that it is an extreme consequence of the normal aging process, that it is dynamic and time-varying, and that it is multidimensional.[26] The nature of these dimensions and whether the multidimensional age-related processes are antecedent to, concurrent with, or a consequence of vulnerability remains unresolved.[27] The temporal and causal relationship[28–30] between the vulnerability of frailty and the adverse outcomes also varies across definitions.[31] A historical examination of the evolution of the construct of frailty may reveal reasons for the coexistence of multiple definitions.

2.1.2 Brief history of frailty as a construct

Frailty, unlike most nosological entities, has a colloquial meaning. Frailty originates from the Latin word *frag-ili-tas*: *frangere* to break, *-ili-* the adjectival form, *-tas* the state. Frailty is literally the state of being easily broken. The common meaning of frailty can describe both the physical characteristics and moral character of a person. The first mention of frailty in the medical literature is often attributed to a 1968 article published in the *British Medical Journal* titled "Old and Frail" describing a shift in the age of patients and highlighting the need for better care organization and provision to older adults.[32] A more formal usage of frailty appears a decade later, in 1979, in demographic studies by Vaupel et al.[33] This line of work aimed to improve mortality modeling by accounting for the heterogeneity in individual susceptibility to all causes of death, i.e., their individual frailty. Although frailty has been used in clinical parlance since the beginning of geriatric medicine, attempts to formalize a definition of frailty were initiated in the 1990s.[25,34] Contemporary formalizations of frailty definitions were made possible by the enrollment of older adults in cohort studies and the development of epidemiologic methods to conduct and analyze data from large cohort studies.[35] The Cardiovascular Health Study[36] and the Canadian Study of Health and Aging,[37] which investigated the epidemiology of cardiovascular diseases and that of dementia, respectively, were the foundation for the development of two objective and parallel conceptual frameworks of frailty:[38] the Fried physical frailty phenotype[39] (PFP) and the Rockwood and Mitnitski deficit-accumulation frailty index[40–42] (FI). Both first described in 2001, they remain the two most influential frameworks of frailty, but have since been joined by more simplified definitions as well as more extensive ones.[7,43] The line between an instrument measuring frailty and a definition of frailty—and between reflective and formative measurement models[31,44]—is thin. The tension between what frailty *is* and how it is *recognized* has been a concern ever since

its introduction in research and clinical practice.[25,27,45] In the last decade, unidimensional and low-dimensional operationalizations of frailty, such as gait speed[46] or the Short Physical Performance Battery,[47] have coexisted with multidimensional,[48,49] latent,[50] and diagnostic-group-related[51] operationalizations of frailty. As a construct, frailty appears to still resist attempts to delineate its common sense meaning into a *single* and well-delimited definition for research and clinical uses.[3]

2.1.3 Levels and potential pathophysiologies of frailty

At its core, frailty seeks to characterize or quantify age-related vulnerability associated with adverse outcomes.[34,52] The specific vulnerabilities and levels of realization[53] considered vary by scale of biological organization from basic molecular mechanisms to cells and organ systems, to individuals, and up to communities and societies.[24,35] There has been a significant interest for frailty across this whole spectrum: at the scales of basic molecular mechanisms to organ systems[54] (i.e., the classical focus of biomedical research) and at the scales of individuals[55] to societies[6,56] (i.e., the focus of epidemiological research). Researchers characterizing frailty at the more fundamental biomedical level have proposed a pathophysiology underlying frailty.

Frailty is understood to be a "disorder of several interrelated physiological systems"[1] which disrupts normal homeostasis.[24,52,57] Under the pathophysiological view of frailty, this disruption of homeostasis is distinguished both from normal aging and from known clinical diseases.[27,58] A decline in the reserve of body systems is expected with normal aging, but the extent of this decline across multiple systems is more pronounced with frailty and accelerates the "normal" incidence of disability. Unlike the typical and simplified pathophysiology of clinical diseases whereby the disease state is brought about by a single converging process (e.g., tumorigenesis for cancer), the pathophysiology of frailty is realized at the level of the underlying biological dysregulation of *multiple* body systems, which in turn cause clinical disease and disability.[57] The antecedence of

frailty to clinical conditions is, however, only partial since most pathophysiological models of frailty involve a feedback loop where clinical disease also affects frailty thus perpetuating and exacerbating it.[57,58] The biological basis for triggering increased susceptibility remains unclear but hypotheses related to DNA damage, mitochondrial dysfunction, telomere shortening, oxidative stress, cell senescence, among others, have been proposed.[24] Specific biological pathways have also been hypothesized for each organ system and are best studied for the brain, the endocrine system, the immune system, and the skeletal muscle.[1,26]

No single level of realization may be inherently correct to posit the pathophysiology of frailty. Although the basic molecular and cell-level mechanisms have proven revolutionary in medical research (e.g., monogenic diseases, infectious diseases), higher scales of analysis may also be explicative[59,60] (e.g., social epidemiology). Of the three conceptual frameworks of frailty discussed below, the Fried frailty phenotype proposes a pathophysiology understanding of frailty, at the organ and individual-functioning levels.[24]

2.1.4 Three major frameworks of frailty

2.1.4.1 The Fried physical frailty phenotype

The Fried physical frailty phenotype[39] is the most widely cited operational framework of frailty.[7] According to Fried et al., frailty is characterized by a specific phenotypic syndrome involving a pathologic "vicious" cycle leading to a progressive decline in physical function and health.[24,61] There are five core components to the phenotype: slowness, weakness, low activity, exhaustion, and unintentional weight loss. The original definition of the first three components is based on distribution-based cut-offs, whereby slowness (measured by gait speed), weakness (measured by grip strength), or low activity are present when individuals are in the lowest quintile of their sex (for all 3 components) and height (for slowness) or body mass index (BMI, for weakness)

subgroup. Exhaustion is defined as feeling "that everything I did was an effort" or that "I could not get going" 3 or more times per week (from the Center for Epidemiologic Studies—Depression Scale[62] [CES-D]). Unintentional weight loss is present when more than 10 pounds was lost in the last year. Together these components capture the vicious cycle of frailty: loss of muscle mass (sarcopenia) leading to a decreased resting metabolic rate and walking speed, leading to reduced activity and total energy expenditure, then leading to chronic undernutrition thus worsening sarcopenia and completing the cycle.[24,61] The frailty phenotype is usually categorized into three frailty levels: individuals with 3 or more criteria have frailty, those with 1 or 2 criteria have prefrailty, and those without any criteria are considered robust or nonfrail. Multiple definitions have been derived from this operational framework. For example, the FRAIL scale uses four of the five criteria (*F*atigue, *R*esistance, *A*mbulation, and *L*oss of Weight) but operationalizes them solely using self-report rather using physical performance tests for slowness and weakness; the presence of greater than 5 *I*llnesses is also substituted for low activity.[63,64] Another example is the Study of Osteoporotic Fractures (SOF) index which uses three of the five Fried criteria: weight loss, inability to rise from a chair five times without using arms, reduced level of energy.[65,66]

2.1.4.2 The Rockwood deficit-accumulation frailty index

The second most cited frailty framework is the Rockwood and Mitnitski deficit-accumulation frailty index (FI).[7,40,41] The deficit-accumulation frailty index conceptualizes frailty as the stochastic process of accumulating age-related deficits which together capture a person's vulnerability to adverse outcomes.[67] The precise nature (and composition) of deficits is not predefined; rather, the proportion of deficits present in an individual is considered to best quantify the level of frailty. Any variable or measurement can be considered as a deficit as long as it conforms to the following four criteria: (i) is related to health status, (ii) increases with age, (iii),

does not saturate at an early chronological age, and (iv) covers multiple systems. An added fifth requirement is to use the same deficits for longitudinal comparisons.[42] Clinical diseases, symptoms, signs, laboratory abnormalities, disability scales, physical performance measurements (e.g., gait speed) have all been used as deficits within implementations of frailty indices. The first frailty index included 92 deficits,[40] a subsequent one 70 deficits.[41] A minimum of 30–35 deficits has been recommended to ensure stability of indices.[42] The composition is not predetermined, and deficits are unweighted in the calculation of FIs; the information contained in the underlying latent variation in deficits takes precedence over the specific informativeness of each included deficit. The frailty index framework has led to multiple implementations using various sources of data such as comprehensive geriatric assessments (FI-CGA[68]), electronic medical records (eFI[69]), laboratory values (FI-LAB[70]), trials[71,72] or registry (mFI-5[73] and mFI-11[74]) data. Across implementations, frailty indices typically range between 0 and 0.70, an empirical level at which frailty is incompatible with survival.[75] Frailty indices can be used both as a continuous variable or categorized into frailty levels, although no consensus cut-off exists.[76]

2.1.4.3 Multidimensional geriatric frailty

The two most influential frameworks of frailty focus on different domains of aging to capture vulnerability. Whereas the Fried phenotype centers on the physical dimension of vulnerability disentangled from disability and multimorbidity,[58] the Rockwood FI centers on the latent vulnerability underlying any age-related health deficits. There is, however, no *a priori* reason to restrict the conceptualization of frailty to the physical domain or to an underlying trait. As long as the notion of vulnerability on multiple dimensions is characterized, some form or idea of frailty can be entertained.[22,26] This may be the basis for a very general conceptualization of frailty as "multidimensional geriatric frailty." This third model of frailty can be considered to encompass all

others of definitions of frailty not derived from the Fried frailty phenotype or from the frailty index. Definitions of frailty falling under this umbrella category exhibit various levels of detail and exhaustiveness. At one end of the spectrum are the Clinical Frailty Scale (CFS)[41,77] and PRISMA-7 (Programme de recherche sur l'Intégration des Services pour le Maintien de l'Autonomie)[78], two instances of synthetic and relatively concise frailty measurements. The CFS quantifies frailty by using a 7-point scale[41] (later expanded to 9 points)[77] based on pictograms and rubrics describing a combination of fitness or physical function level, medical conditions, disabilities, and life expectancy. PRISMA-7 includes seven binary items assessing chronological age, sex, health problems limiting activity, need of assistance on a regular basis, health problems requiring home stay, social support, and use of mobility aids. At the other end of the spectrum are the Edmonton Frail Scale (EFS)[48] and the Tilburg Frailty Indicator (TFI)[49], two instances of frailty measurements with exhaustive coverage of geriatric domains. The EFS assesses nine domains (cognition, general health status, functional independence, social support, medical use, nutrition, mood, continence, and functional performance) using 11 questions for a total score of 17 points. The TFI assesses three domains (physical, psychological, and social) using 15 binary questions.

2.1.5 A brief summarizing comparison of frailty frameworks and definitions

Due to the vast numbers of definitions currently in usage, detailing every frailty definition would be unwieldy.[7,8,43] One way of summarizing frailty definitions is to relate them to one of the three frameworks described above. Another useful way is to summarize definitions, and their overarching frameworks, ontologically, i.e., by the concepts, categories, and relations each definition entails. By no means exhaustive, Table 2.1 suggests nine dimensions by which frailty definitions can be distinguished.

Table 2.1. Dimensions and comparisons between frailty frameworks and definitions

DIMENSIONS	DISTINCTIONS	EXAMPLES
Conceptualization of pathology	*Process* pathophysiological syndrome with etiological contribution	Fried frailty phenotype FRAIL scale
	State of vulnerability	Rockwood frailty index Modified frailty indices
Domain coverage	Focus on physical frailty (e.g., exclusion of cognition or social support)	Fried frailty phenotype FRAIL scale
	Variable	Rockwood frailty index and derived implementations
	Explicitly wide	Edmonton frail scale Tilburg Frailty Indicator PRISMA-7
	Implicitly wide	Clinical frailty scale
Distinction from other age-related concepts	Cognitive impairment	*Yes*: Fried frailty phenotype *No*: Tilburg Frailty Indicator
	From disability and chronic conditions	*Yes*: Fried frailty phenotype *No*: Rockwood frailty index, FRAIL scale
	From psychological impairment and social support	*Yes*: FRAIL scale *No*: Tilburg Frailty Indicator
Measurement model	Formative	Rockwood frailty index Modified frailty indices
	Reflective	Fried frailty phenotype PRISMA-7
Source of data and assessment	Solely self-report	FRAIL scale Tilburg Frailty Indicator
	Combination of self-report and objective measures	Fried frailty phenotype Edmonton frail scale
	Variable	Rockwood frailty index
Level of detail of assessment	Brief	Clinical frailty scale PRISMA-7
	Detailed	Rockwood frailty index (70-item) Edmonton frail scale
Frailty variable type	Categorical	Fried frailty phenotype FRAIL scale
	Discrete	Clinical frailty scale
	Continuous (can also be categorized)	Rockwood frailty index
Method to establish frailty cut-offs	Distribution-based cut-offs for components or resulting frailty measurement	Rockwood frailty index
	Absolute cut-offs	PRISMA-7 Clinical frailty scale Edmonton frail scale

2.1.6 Importance and influence of frailty

The many unresolved definitional questions about frailty have not impeded its uptake among researchers in aging. Frailty has been the focus of considerable research interest, and calls for its implementation in clinical practice have also been made in recent years.[9–11,79] Premised on the strong and consistent associations between frailty and adverse outcomes (e.g., institutionalization, disability, functional decline, and death), calls for the incorporation of frailty into clinical practice have suggested routine screening for frailty, interventions to prevent or reverse frailty, using frailty to alter treatment indications, or modify specific interventions. Both within and outside geriatric medicine, frailty is a thriving area of research and has made its way into original research, reviews, and clinical practice guidelines[80,81] across multiple disciplines such as cardiology,[82] hepatology,[83] infectious diseases (e.g., HIV),[84] nephrology,[85] oncology,[55] pulmonary medicine,[86] and surgery.[87,88] Calls for incorporation appeared to remain mostly unheeded; to what extent the incorporating of frailty to practice has led to improve patient outcomes has not been well explored.[11]

2.2 Aging, geriatrics, and syndromic approach to complexity

A full understanding of any topic hinges on considering the backdrop on which it stands, i.e., the scaffolding "under" the genesis of frailty. The construct of frailty and its development are intertwined with the demographic transition and the development of geriatric medicine and age-specific nosological constructs.

2.2.1 Aging and geriatric medicine

The demographic transition that industrial societies have undergone in the twentieth century, with high fertility rates that were followed by lower rates along with improvement in life expectancy, has increased both the proportion and absolute numbers of Canadians 65 years and older. From

1971 to 2020, the proportion of older adults grew from 8% to 18% of the general population to reach more than 6.8 million currently over 65 years old.[89] Meeting health care needs of its aging population is one of the major challenges facing the Canadian health system.[79] Although this challenge is more acute than ever, it is not novel. The birth of geriatrics as a medical discipline is often attributed to Ignatz Leo Nascher and his coining of the word "geriatrics" in a 1909 article:[90]

> *Geriatrics, from* geras*, old age, and* iatrikos*, relating to the physician, is a term I would suggest as an addition to our vocabulary, (…) to emphasize the necessity of considering senility and its disease apart from maturity and to assign it to a separate place in medicine.*

In 1914, Nascher published the seminal book *Geriatrics: The Diseases of Old Age and Their Treatment* in which he recognizes the inevitable difficulty in classifying disease in old age due to the interrelations "between senility apart maturity and its diseases, as sui generis senile diseases" (p. 64).[91] Noting that "[t]here is probably no other branch of science in which nomenclature and classification are as imperfect as in medicine," he categorizes diseases found in the aged into five categories: "(i) Primary senile diseases (i.e., diseases in which there is an increase, decrease or perversion of the ordinary […] senile changes); (ii) Secondary senile diseases, i.e., disease which results from the senile change; (iii) Modified disease of old age, i.e., diseases which, when occurring in old age are modified by the senile conditions […]; (iv) Preferential diseases of old age, i.e., diseases which occur most frequently in advanced life; (v) Diseases uninfluenced by age or are rare in old age" (p. 65). More than 100 years since this classification, Nascher's initial forewarning that "a revolutionary revision of our nomenclature is necessary before we can place upon a scientific basis medical terms and the classification of disease, and until this is accomplished every classification must be imperfect" still holds true. In the last five decades of research and practice in aging, geriatrics has developed its own disease classification to face the

complexity in the health states of old age. The description of frailty and the pursuit of a formal definition of the frailty construct proceed from this thread.

2.2.2 Complexity and geriatric syndromes

The complexity of geriatric medicine takes root in its full consideration of a life stage.[92] Disciplines of medicine are prototypically concerned with either an organ system (e.g., internal medicine and subdisciplines) or an anatomical site (e.g., surgery and subdisciplines). Not unlike the case of ped*iatrics* or psych*iatry*, the essence of geriatrics is not realized at an organ system or anatomical level. Old age health is influenced by the interrelations in the multiple domains of functional capacity (activities of daily living), cognition, mobility, mental health, social support, and living environment, in addition to usual medical conditions.[93] The prevailing epistemology of medicine is one of reductionism in causal explanations. Single causes—allowing for potential interactions—underlie pathological states[94,95] as best exemplified by acute infectious diseases. Not all pathological states, however, are conducive to reduction to a primary cause,[96] especially when they are taken to include impairments in one or many age-related domains; impairments in activities of daily living, in cognition, or in mobility are not subsumable under standard medical mechanistic causality. A novel overarching nosological entity progressively emerged in geriatric medicine starting with Bernard Isaacs' four I's and Geriatric Giants in 1965:[97] immobility, intellectual impairment, and incontinence were the precursor to the now-canonical *geriatric syndromes*.[17,18] Although the precise categorization of conditions as geriatric syndromes varies, dementia (or cognitive impairment), depression, delirium, functional impairment, malnutrition, incontinence, vertigo and syncope, falls and impaired mobility, elder mistreatment, polypharmacy, and pressure ulcers are instances of geriatric syndromes.[18,98,99] Frailty is also considered by many as a foundational geriatric syndrome.[9,61,100] Unlike the classical medical syndrome whereby a single

pathophysiological process leads to protean manifestations (e.g., Cushing's syndrome), geriatric syndromes flip the "etiological" chain: multifactorial causes and accumulated effects of multiple impairments lead to a single unified manifestation. The necessary criteria of geriatric syndromes are a single phenomenology (or phenotype), commonness with age, multifactorial etiology, impairment associated with a worse outcome.[17,101–103]

The geriatric-syndrome approach to diseased health states of old age may be a useful framework to capture the complexity of aging.[17] At a minimum, it draws attention to the importance of certain phenomena of aging, such as delirium, guarding them from omission or from irrelevance to medical consideration. Geriatric syndromes can also accommodate multiple causes—and their potential interactions—in the etiology and explanation of health states. However, the adaptability of geriatric syndromes may also have a cost. Rather than requiring a causal explanation of disease (which also entertains the idea of reversibility or intervenability), geriatric syndromes may be constructed on the sole basis of risk factors and an association with worse outcomes. This low ontological bar may lead to abdication, if not rejection, of the expectation that a diagnosis must relate to identifying a cause, which leads to intervenability and reversibility of the many causes underlying geriatric syndromes (and, as a consequence, of the geriatric syndromes themselves). For many geriatricians, "diagnostic strategies to identify the underlying causes [of geriatric syndromes] can sometimes be ineffective, burdensome, dangerous, and costly."[17] Moreover, "therapeutic management of the clinical manifestations can be helpful even in the absence of a firm diagnosis or clarification of the underlying cause."[17] Whether Occam's razor or Hickam's dictum[104] holds most wisdom in geriatric medicine is still undecided. One thing is clear: the construct of frailty emerged amid the backdrop of geriatric syndromes as a nosological development in response to the complexity of aging. Whether this response has proven

fruitful likely varies by the lens used to examine the construct of frailty. How should frailty be apprehended: under the scope of epidemiology, a study, or the scope of medicine, a practice?

2.3 EPIDEMIOLOGY, MEDICINE, CLINICAL PRACTICE, AND FRAILTY

2.3.1 Epidemiology

Epidemiology is the study and analysis of the distribution, patterns, and determinants of health and disease conditions in defined populations. Under the lens of epidemiology, whether frailty is a *health* condition or *disease* condition does not alter its relevance as a topic of study. The distribution, patterns, and determinants of frailty, as the outcome, can be studied and analyzed. In addition to being an outcome under study, frailty can also be investigated as a *determinant* of health and disease conditions and as a *defined* population. Association and inferential epidemiological studies interrelate populations, interventions (or exposure, "determinant"), comparators, and outcomes (PICO),[105] usually using quantitative methods to draw relations between interventions and outcomes.[106] Observational studies may further include frailty as an additional variable to PICO (e.g., as a confounder: stratification variable, covariate in regression analyses, or propensity scoring.[107–109] The current literature on frailty comprises studies where the frailty construct is used as one of these epidemiological functions. Though calls for consequentialist epidemiology[110] have been made and the limits of risk factor epidemiology raised,[111] the methods of epidemiology are permissive of constructs as objects of study. Geriatric syndromes, such as frailty, delirium, falls, incontinence, cognitive impairment, can be studied under the methods (and constraints) of epidemiology. Studies on these established geriatric entities have proliferated. Expanded data availability and analysis methods have further sparked the development, adoption, and study of further geriatric nosological constructs using large datasets and contemporary methods: motoric cognitive risk syndrome[112] (combination of cognitive

impairment and slow gait), multimorbidity[12,113] (clustering of chronic conditions), biological age,[114] among others. Whether interest in these entities—old and new, with frailty chiefly among them—will prove durable may well depend on their ability to be translated into action. Whether objects of epidemiological study will stand the test of time will depend on their ability to prove applicable and useful for practitioners of public health for epidemiology, and of medicine, for clinical epidemiology specifically.

2.3.2 Medicine and clinical practice

The lens of medicine has been continuously polished since the days of Hippocrates. The scope of medicine, what is considered under its purview, has been subject to constant revision, most of it tacit and progressive.[115–119] Western modern medicine remains under the predominance of the biomedical model tradition: causal and pathophysiological disruptions of biological processes lead to abnormal human functioning, i.e., disease.[19,120] The locus of disruption may have varied in the last centuries and decades from infectious agents, monogenetic and polygenic defects, inflammation and degeneration, to microbiota, but a constant endures: ultimately, this disruption should eventually lead to intervention. Medical training and the prevailing conceptualization of clinical reasoning are strongly premised on uncovering, preventing, and intervening on the unifying cause and alleviating its consequence when and if possible.[121] Yet, as previously noted, the biomedical model, although predominant, may be insufficient to subsume the workings of less prototypical medical disciplines, such as psychiatry, addiction medicine, palliative medicine, or geriatrics. Moreover, the classic biomedical model does not fully account for the much more complex organization of knowledge in clinical decision-making and processes.[122–124] For example, study eligibility criteria, vital signs, risk prediction, or prognostication do not naturally fit the biomedical etiologic view of medicine. Relying only on the lens of biomedical medicine to infer

the clinical practice potency or applicability of frailty may be insufficient. There is a plethora of theoretical constructs both from research and practice; of those, some that are translated into clinical processes, usage, and practice. How does a construct move from theory to the practice? At a minimum, practice imposes *clinical* and *individual* actions : a successful nosological entity must alter a clinical action for individuals.

2.4 MALLEABLE CATEGORIZATIONS, DEFINITION AND FUNCTIONS OF CONSTRUCTS

Reconciling the many forms of frailty has been the focus of intense debates in geriatric research in the last two decades.[25,38,125,126] No consensus has been reached yet,[3] and it may be illusory to reach a unique understanding for frailty if many diverging purposes for its existence live side-by-side. Naming or measuring frailty is done for a purpose; the variability in functions of frailty, some of which are germane to epidemiology and others to clinical practice, may hold the key to resolving part of the ongoing definitional (and ontological) debate. Within the limits that age-related frailty identifies humans or some characteristics of them, “frail” and “frailty,” as an adjective and a substantive, can predicate or be predicated by any given concept.[127] The malleability of frailty may have led to unrecognized confusion about what it is (its definition), what it covers (its substance/ontology), and what it is for (its function/purpose). As “truth emerges more readily from error than from confusion,”[128] it may be preferable to mistakenly, but deliberately, attempt to determine the many functions of frailty, than to further compound the confusion by proposing yet another definition for it. In clinical practice and reasoning, the definitions may not be fully distinguishable from their functions. No formal and integrative mapping of nosological functions has been conducted; this may be an impossible program, as it may mean seeking to formalize epidemiological and clinical language and content themselves.[119,127] I would suggest, however, that all things (i.e., frailty and other constructs) defined at the intersection of epidemiology and

clinical medicine should fulfill the following three general minimum criteria to enable any eventual purposeful use:

1. **CATEGORIZATION AND IDENTIFICATION**. Categorization of persons or of features that can characterize and identify persons and thus the implicit *distinction or contrast* between persons. This categorization can include interventions done to persons.
2. **HEALTH RELATION.** Direct or eventual relation with health states and disease conditions.
3. **FUNCTIONAL CLASS MEMBERSHIP.** Interrelations with other constructs in-use in discipline-accepted ways, as the following examples may clear. These functional classes may be particular to:
    a. **Epidemiology**, of which classes such as population, intervention, comparator, outcome, and confounders are examples.
    b. **Medicine (and its specific disciplines)**, of which classes such as symptom, sign, risk factor, diagnosis, and prognosis are examples.

This third criterion is very broad as the focus and methods of both epidemiology and medicine are wont to evolve. Beyond these three proposed criteria, in the context of *clinical* epidemiology, an additional criterion may be added related to the *usefulness* of a construct when in the context of *clinical practice*. As previously mentioned, practice implies some form of action:

4. **ACTION RELATION.** Direct or eventual relation to altering (i.e., indicate or preclude) an action on persons, specifically due to this construct.

Together, these four criteria may constitute a foundation to examine and analyze any nosological construct proposed or in usage. The research presented in this dissertation builds on this foundation to examine frailty: its basis and its potential for use in clinical medicine.

2.5 Overarching goal: frailty, constructs, and the clinical imperative of medicine

Not all nosological constructs are equally useful. Not all categorizations have an equal importance in medical practice. Medicine is "the science and the practice of caring for a patient and managing the diagnosis, prognosis, prevention, treatment or palliation of their injury or disease."[129] Although the scope of medicine is large and encompasses both research (i.e., "science") and practice, ultimately, for clinicians (and epidemiologists) the reward of the work is the improvement of individual and populational health. The "science" of medicine and the constructs it features must reach the "practice" of medicine to affect health. Ensuring that categorizations studied and described in research eventually make a clinical difference on health outcomes is a crucial goal.

Despite the mounting interest in frailty, it remains unclear whether it has or will materially make a difference to the health outcomes of older adults.[11] Taking the latter as the *clinical imperative of medicine*, the overarching goal of this dissertation is to explore the conceptual underpinning and clinical applicability of age-related nosological constructs, with an emphasis on frailty. More specifically, four facets of the coupling between construct and applicability will be examined. The first two questions pertain to categorization and identification. First, whether there exists a substrate to specifically categorize older adults as frail (i.e., heterogeneity in aging as a *substrate for categorization*). Second, whether a widely used measure of frailty can categorize older adults in a consistent manner (i.e., reliability of frailty as a *stable and usable categorization*). The third question examines the criteria on which studies and findings about constructs may be considered clinically applicable as based on their *health relation, functional class,* and *action relation*. The fourth and final question will examine whether recent frailty research has indeed proven clinically applicable.

At the crossroads of epidemiology and clinical medicine, this dissertation uses on interdisciplinary and pragmatic methods to bridge research concepts and eventual clinical actions. In asking whether frailty can be translated into practice—and in aiming to identify the underlying reasons—this dissertation engages with both the *substantive content* of frailty and the *methods used to understand* frailty. More generally, as will be discussed, the research conducted aims to be easily generalizable to better understand the requirements for any construct to be clinically "successful," i.e., fulfilling the clinical imperative of medicine. At its essence, this dissertation is about nosological constructs in geriatric medicine, their usefulness, and how this usefulness might be determined.

Chapter 3. Data source and methods

3.1 DATA SOURCE

3.1.1 Canadian Longitudinal Study on Aging cohort

Manuscripts 1 (Chapter 4) and 2 (Chapter 5) use data from the Canadian Longitudinal Study on Aging (CLSA). The CLSA is a population-based and nationally stratified study of 51,338 Canadian residents aged between 45 and 85 years at the time of recruitment.[130,131] Participants were enrolled between 2012 and 2015, at baseline, into one of two cohorts: the Tracking cohort (n = 21,241) and the Comprehensive cohort (n = 30,097). Measures were collected using 60- to 70-minute computer-assisted telephone interviews for the Tracking cohort. For the Comprehensive cohort, participants were assessed by in-person interviews at home and by additional questionnaires, tests, performance measurements, and biological specimens at one of 11 data collection sites. The data were obtained through the CLSA Data Access process.

3.1.1.1 Eligibility criteria

Canadian residents aged 45 to 85 years old were eligible for recruitment. Exclusion criteria were being unable to read and speak in either French or English, living in long-term care institutions (i.e., those providing 24-hour nursing care), presence of cognitive impairment at the time of recruitment as determined by CLSA interviewers, residents in the three Canadian territories, persons living on federal First Nations reserves, and full-time members of the Canadian Armed Forces.

3.1.1.2 Measures

The CLSA collected data on an extensive set of age-related domains including social and demographic measures, health status, physical performance measurements, psychological measures, behavioral measures, and health care utilization. The Comprehensive cohort included

specific cognitive measures about memory and executive function, medications, physical measures (i.e., anthropometric, physical function, vision, and hearing), electrocardiogram, lung function testing, and blood and urine specimens. For participants who consented, data linkage was done with administrative health databases for death ascertainment in some provinces. The CLSA updated the mortality data in July 2019. The mortality data came from the next of kin contacting CLSA directly, identification of death at the time of follow-up, and from linkage to provincial vital statistics. Because date of death was not available, mortality status was determined for all participants in July 2019.

3.1.2 Analytical samples for Manuscripts 1 and 2

Manuscript 1 explores changes in variability in health characteristics by age groups. The analytical sample included the full CLSA Comprehensive cohort (n = 30,097 participants). Manuscript 2 examines the reliability of frailty measures when used in the older adult population. As frailty is an age-related concept and is typically measured in older adult populations, the analytical sample was restricted to adults 65 years and older with mortality data as of July 2019 (n = 12,080 participants).

3.2 ETHICAL APPROVAL

The studies in Manuscripts 1 and 2 were approved by the McGill Faculty of Medicine and Health Sciences' Institutional Review Board.

3.3 OVERVIEW OF METHODS

3.3.1 Manuscript 1 – Health heterogeneity in older adults: exploration in the Canadian Longitudinal Study on Aging

In its commonsense meaning, *heterogeneity* is the quality or state of being diverse in character or content. Diversity can be determined in many ways and about many different things; I used a

statistical measure of diversity based on distance from a central location of 34 health-related variables. Manuscript 1 explores the *change* in heterogeneity as function of chronological age. For each variable, the respective distance from the central location was measured as the deviation from the mean for each 1-year age and sex group; the means were predicted using a locally estimated scatterplot smoothing regression for each age and sex group. Each participant has an absolute deviation value for each variable from their age and sex group; this deviation was then regressed on chronological age to determine whether there is an association. When an association was present, I examined whether the magnitude was both statistically and clinically significant. For significant associations, two additional measurements issues were examined that could explain the age and deviation relationship: (i) a mean-variance relationship and (ii) the scaling of the variable. First, the variance (and deviation) of many variables may increase with their mean; to adjust for this relationship, I examined whether chronological age was associated with variables (with their mean). When there was an association, the mean value of the variable by age-sex group was used as an adjustment variable in the regressions of deviation on chronological age. Second, variables used in research in aging and clinical practice do not all represent natural phenomena measured on an objective scale (e.g., blood pressure). Many quantitative variables in geriatrics are designed (i.e., scaled) to be clinically differentiating (normative) and informative which may affect the deviation as a function of chronological age. The presence of normative-clinical scaling was determined by consensus (QDN, MFF, PD) for each variable.

In addition to examining each variable individually, I examined the heterogeneity of the eight health domains encompassing the 34 variables. A measure of multivariate variability was used as the dependent variable for regression on chronological age. I used effective variance which

is a dimensionless measure which quantifies the average (and standardized) multivariate scatter of variables.[132,133]

In Manuscript 1, I also investigated the importance of variability due to heterogeneity *within* the oldest age group (or *between* participants 85 years old) relative to the difference between the oldest age group compared to younger age groups. For each variable I first estimated the average absolute deviation of participants 85 years compared to their age and sex group. I then regressed the variable of interest on chronological age: the age coefficient represents the average difference *between* age groups. To compare those two quantities, I identified the age at which the difference between age groups became greater than the difference within the age group of 85-year-olds.

3.3.2 Manuscript 2 – Clinical correlates and implications of the reliability of the frailty index in the Canadian Longitudinal Study on Aging

Manuscript 2 investigates the reliability of an important frailty measurement framework: the frailty deficit-accumulation index. Reliability is measured using standard statistics such as intraclass correlations coefficients or standard errors of measurement that contrast the variance due to measurement versus that due to individuals.[134,135] In investigating the reliability of frailty, I used these standard methods but also supplemented them with methods that provide more clinically oriented estimates of reliability. I used Monte Carlo methods to implement various frailty indices varying in composition of items and number of items and age-related domains. Each frailty index was derived and examined in the same CLSA analytical sample for clinically relevant estimates: frailty scores, the prevalence of frailty, cut-offs of frailty, odds ratios for mortality, and predicted risk of mortality. The combination of standard statistics and clinically anchored estimates provides

a comprehensive and interpretable understanding of the reliability—and thus clinical applicability—of frailty indices.

3.3.3 Manuscript 3 – Appraising clinical applicability of studies: mapping and synthesis of current frameworks, and proposal of the FrACAS framework and VICORT Checklist

Manuscript 3 used a systematic search strategy and qualitative analysis. I searched the published literature (inception of PubMed and EMBASE to November 2020) for articles presenting frameworks and criteria to appraise the clinical applicability of studies. Selected articles were identified in full to identify unique frameworks. I performed conceptual thematic analysis to iteratively map frameworks and to synthesize criteria of applicability. To complement existing frameworks and to emphasize *clinical* applicability, I developed and proposed a novel framework for appraisal by integrating four major inputs: contemporary debates in epidemiology and clinical research, brainstorming and discussion meetings with methodologists and clinicians, comparison with existing frameworks, and pilot application of iterative versions of the framework in a scoping review on clinical frailty (Manuscript 4). Meetings involved interdisciplinary substantive and methodological contributions. Preliminary versions of the framework were iteratively tested and refined to reach the final consensus framework.

3.3.4 Manuscript 4 – The State of frailty in research: a mapping review of its clinical applicability to practice

Manuscript 4 sought to examine whether frailty studies are applicable and translatable to clinical practice. Due to the prohibitive number of articles published on frailty, I used a systematic sampling procedure to select articles to appraise. All articles published in 2017 or 2018 on the topic of clinical (i.e., geriatric) frailty were identified. The articles were randomly ordered, and the order was followed to appraise articles until 150 articles were identified that used frailty in a

function that could inform clinical practice: predictor, mediator, effect modifier, or primary selection criterion. We enumerated all other functions of frailty (e.g., outcome measure or confounder) in the studies appraised to reach the 150 where frailty was used in a practice informative function (FUNCTIONAL CLASS). For these 150 fully appraised articles, I extracted data on the operationalization and specific cut-offs used for frailty (CATEGORIZATION). I then adapted and used the appraisal framework developed in Manuscript 3 to classify studies by their potential for clinical usefulness and applicability (ACTION RELATION): practice-changing (and informing), practice informing (but not practice changing), or nor practice informing.

Chapter 4. Manuscript 1: Health heterogeneity in older adults: exploration in the Canadian Longitudinal Study on Aging

4.1 PREFACE

Research specific to aging and geriatric medicine presumes that older adults are different from younger adults. Differences between older and younger adults form the essential foundation for age-related constructs of which frailty is an important exemplar. Age differences may be due to differences between the "average" older adult and the "average" younger adult, but this difference may be compounded or overshadowed by the differences between an individual older adult compared to the "average" older adult, that is, by the differences between older adults themselves. The latter difference has been referred to as heterogeneity in aging, a central dictum of research on aging. This heterogeneity is what age-related constructs attempt to better capture, characterize, and quantify. The overarching goal for this dissertation is to explore the conceptual underpinning and clinical applicability of age-related nosological constructs, with an emphasis on frailty. In keeping with this goal, the research presented in the first manuscript investigated whether there is indeed greater heterogeneity with chronological age, i.e., whether there is a *suitable substrate* for categorizing older adults differentially from their younger counterparts, whether the health characteristics and health domains showcasing greater variability can be identified, and reasons for the variation in heterogeneity.

***

This manuscript has been published in the Journal of the American Geriatrics Society with an accompanying editorial by Luigi Ferrucci and George A. Kuchel (2021, Heterogeneity of Aging: Individual Risk Factors, Mechanisms, Patient Priorities, and Outcomes. J Am Geriatr Soc, 69: 610–612).

4.2 TITLE PAGE

Title

Health heterogeneity in older adults: exploration in the Canadian Longitudinal Study on Aging

Short title

Exploring health heterogeneity in aging

Authors and affiliations

Quoc Dinh Nguyen, MD, MA, MPH,[a,b,c] Erica M. Moodie, PhD,[c] Marie-France Forget, MD, MSc,[a] Philippe Desmarais, MD, MHSc,[a,b] Mark R. Keezer, MDCM, PhD,[b,d] Christina Wolfson, PhD[c,e,f]

a. Division of Geriatrics, Centre hospitalier de l'Université de Montréal, Canada;
b. Centre de recherche du Centre hospitalier de l'Université de Montréal, Canada;
c. Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Canada;
d. Departments of Neurosciences & Social and Preventative Medicine, Université de Montréal, Canada
e. Department of Medicine, McGill University, Canada;
f. Neuroepidemiology Research Unit, Research Institute of the McGill University Health Centre, Canada.

Corresponding author

Quoc Dinh Nguyen
1000, Saint-Denis, Montreal, Quebec, Canada H2X0C1
quoc.nguyen@mail.mcgill.ca

Word counts

Abstract: 272
Manuscript Body: 3364
Tables: 3
Figures: 2
References: 35

4.3 ABSTRACT

Background

A widely held dictum in aging research is that heterogeneity in health increases with age, but the basis for this claim has not been fully investigated. We examined heterogeneity at different ages across health characteristics to describe variation and trends; we investigated the comparative importance of between-age versus within-age heterogeneity.

Design

Cohort study

Setting

Community-dwelling older adults

Participants

30,097 adults aged 45 to 86 years from the Canadian Longitudinal Study on Aging

Measures

34 health characteristics in 8 domains (physical measures, vital signs, physiological measures, physical performance, function/disability, chronic conditions, frailty, laboratory values) were assessed cross-sectionally. We used regression models to examine heterogeneity in health characteristics (using absolute deviation) and domains (using effective variance) in relation to age. Comparison between between-age and within-age heterogeneity was quantified by estimating the age threshold at which the former exceeds the latter.

Results

Of 34 health characteristics, 17 showed increased heterogeneity, 8 decreased, and 9 no association with age. The associations between heterogeneity and age increased, generally, but were nonlinear

for most domains and non-monotonic for some. We observed peak heterogeneity at approximately 70 years. Between-age heterogeneity, compared to within-age heterogeneity, was most important for forced expiratory volume in 1 second and grip strength, but varied across characteristics.

Conclusion

Overall health heterogeneity increases with age but does not uniformly increase across all variables and domains. Heterogeneity in aging reinforces the need for geriatric assessment and personalized care, depends on which health characteristics are assessed, their measurement properties, and their referent group. Our findings suggest further research to develop improved single-dimension and multidimensional instruments, as well as specific vital and laboratory reference ranges for older adults.

Keywords

CLSA, measurement, variability, heterogeneity

4.4 INTRODUCTION

Aging is not uniform. Older adults have variable health states, embodied in different levels of functioning,[1] chronic conditions, and mortality rates.[2] A widely held dictum in aging research is that older adults show greater heterogeneity in health metrics (or "health heterogeneity") than their younger peers.[3,4] Heterogeneity can be defined as the quality or state of being diverse in character or content.[5] Although older adults differ from younger adults, heterogeneity in aging suggests that older adults would be increasingly different among themselves as they age, quantified as progressively greater deviation (spread) from the average value of their age group.

The ongoing search for a better understanding of heterogeneity[6] has provided the impetus for the development of age-related constructs such as multimorbidity, disability, and frailty.[1,7–11] Variability is a fundamental tenet in clinical practice where it underpins reference ranges for vital signs and laboratory tests. However, the central premise that heterogeneity increases with chronological age appears to have been assumed rather than substantiated.[3] Heterogeneity has been often alluded to and reported inconsistently in the literature in the last fifty years.[6,12–17] But it has not been the focus of deliberate empirical research, in particular in contemporary populations.[4,18,19] In the absence of research, increasing heterogeneity as adults age may not pertain to all health characteristics, may not have a consistent trajectory, and may require consideration of the measurement scales used.[3,15]

In this study using data from the Canadian Longitudinal Study on Aging (CLSA), our first objective was to explore whether, and how, heterogeneity increases in individuals from middle to advanced age. We examined heterogeneity in health characteristics and domains to describe variation and trajectories and explanations for changes in heterogeneity with aging. Our second objective was to compare the heterogeneity due to differences between age groups with the

heterogeneity due to differences within age groups. The overarching objective of this study was to explore how features of heterogeneity in older adults, as a central tenet of aging, might inform geriatric practice and research.

4.5 METHODS

4.5.1 Cohort

We used cross-sectional data from the CLSA which enrolled a nationally representative sample of over 51,000 participants aged 45 to 85 years, at baseline, into an in-person assessed cohort (Comprehensive) or a telephone-only assessed cohort.[20,21] Because data for physical assessments were required, we used the Comprehensive cohort comprising 30,097 adults recruited and assessed from 2012 to 2015. Exclusion criteria for the CLSA were people living in the Canadian territories, on a First Nations reserve, or in institutions, being full-time members of the Armed Forces, having cognitive impairment (as assessed by interviewers), and being unable to respond in French or English.

4.5.2 Variables, health characteristics, and domains

We examined the heterogeneity of 34 health characteristics in 8 domains (physical measures, vital signs, physiological measures, physical performance, function and disability measures, chronic conditions, frailty, and laboratory values), summarized in Supplemental Methods 4.1 (and reported in Table 4.1). These characteristics were chosen based on their usage as heterogeneity-describing variables used in research or clinical practice. Physical measures, vital signs, physical performance, physiological measures, and laboratory values were assessed in-person. Function, disability, and chronic conditions were self-reported; for details, see the CLSA Protocol.[21] Based on a standard procedure previously described[22] we derived a frailty index using 34 variables as reported in the Supplemental Methods 4.1.

4.5.3 Analysis

Heterogeneity by health characteristics and domains

We used boxplots by 10-year age bins and sex to examine the heterogeneity (spread) of each variable. The quantitative association between heterogeneity and age was evaluated in three sequential steps for each health characteristic. First, we derived a variable representing an individual's absolute deviation from their predicted age-sex-group mean value; this absolute deviation was regressed on age (1-year bins) and sex (Model 1). We assessed statistical and clinical significance of the age coefficient to determine whether heterogeneity increases with age. Clinical significance was defined as a change in deviation over 40 years (age coefficient*40) that was greater than the last clinically relevant unit of a measurement (e.g., 1 kg for weight, 0.01 m for height). Second, as some health characteristics may have a mean-variance relationship (i.e., increasing variance as the mean increases), we examined the relationship between the *mean* of a variable and age in linear regression models adjusted for sex (Model 2). If the variation in mean was in the same direction as the variation in deviation, we further adjusted Model 1 for this mean-variance relationship by adding the age-and-sex-specific mean as a covariate (Model 3). If the age coefficient changed direction or was no longer clinically or statistically significant, we considered that the change in heterogeneity was explainable by a mean-variance relationship. Third, as the scaling of a measure may influence heterogeneity, we determined whether the health characteristic had a normative or clinical scaling through consensus (QDN, MFF, PD). Normative-clinical scaling was present when a measure had a floor or ceiling effect, *or* when the range of measurement and scaling of a characteristic was determined by clinical purpose. See Supplementary Figure 4.1 for a flowchart and descriptions of analytic steps and clinical significance.

To describe heterogeneity by health domains, we used effective variance, a summary measure of heterogeneity for multiple variables, which quantifies the average multivariate scatter of variables.[23] After stratifying by sex, we normalized all variables to ensure equal weighting when computing effective variance for each 1-year bin. Linear regression was used with effective variance regressed on age and sex. This was done for each domain, all domains, and all domains excluding laboratory values. To detect nonlinear relationships with age, we tested for quadratic and cubic terms. We retained statistically significant terms to plot predicted effective variance by age and sex.

Heterogeneity between age-group versus within-group

Finally, to determine whether heterogeneity was attributable to variation in age, we compared heterogeneity *between* age groups to that *within* the older age group (i.e., between individuals of that age group), using age 85 years as the reference group; the older age group boundary being more relevant to gerontology and geriatric medicine. For each health characteristic, we predicted: (i) the individual differences, as the average absolute deviation of individuals aged 85 years from their predicted sex-group mean value, and (ii) computed the linear coefficient of age predicting that characteristic. We then used this model to determine the age at which heterogeneity due to variation in chronological age exceeds that due to individual differences. The *closer* the age threshold is to 85 years, the more influential the between-age-group variation is compared to within-group variation. To avoid extrapolating beyond available data, we only report results where this age was between 45 and 85 years. Confidence intervals were computed via bootstrapping. Supplementary Figure 4.2 provides an explanation of the calculation.

Analyses were performed using R 3.6.1 (R Foundation). As our objectives were exploratory, we used available case analyses for variables with missing data, reported in Supplementary Table 4.1, without sampling weights.

4.6 Results

4.6.1 Cohort description and mean variation in health characteristics

Our study sample comprised 30,097 participants of which 15,320 were women (50.1%). The mean age was 63.0 years (SD:10.2; range 45–86 years [all enrolled at age 85 years or younger]), and age groups were well represented. Table 4.1 describes all health characteristics examined. Impairment in activities of daily living (ADL, 2.8%) and instrumental ADL (iADL, 5.2%) was infrequent. Overall, the mean frailty index was 0.09. Hypertension (37.0%) and arthritis (26.4%) were the most prevalent reported chronic conditions. Most health characteristics across all domains showed mean changes with age, especially in participants aged 75–86 years. Older participants had lower mean values of physiological, physical performance, and functional measures. Chronic conditions accumulated with age (mean count of 0.9 in participants 45–54 years to 2.3 in those 75–86 years). The mean frailty index increased from 0.05 (0.06) in those aged 45–54 years to 0.14 (0.09) in those aged 75–86 years.

4.6.2 Heterogeneity by health characteristics and domains

Boxplots for all variables by 10-year age group are reported in Supplementary Figures 3, and standard deviations of health characteristics by age group are reported in Table 4.1. When assessed qualitatively by quantile differences and SDs, the spread of many variables varied with age. Table 4.2 reports the linear relationship between heterogeneity and chronological age for each health characteristic. Of the 34 variables examined, 17 showed clinically significant increased heterogeneity, 8 showed decreased heterogeneity, and 9 no evidence of an association (Table 4.2,

first column). By domains, physical measures (weight, BMI, and waist circumference) showed decreasing heterogeneity with age. Conversely, number of chronic conditions and frailty index showed increasing heterogeneity. Within physiological, physical performance, and functional measures, heterogeneity showed diverging associations with age. For example, heterogeneity in grip strength, Physical activity scale for the elderly (PASE), and forced expiratory volume in 1 second (FEV1) decreased but increased in bone mineral density, Timed up and go (TUG), and Life Space Assessment (LSA). Of the 25 variables with clinically significant associations, 11 (8 with increasing heterogeneity, 3 decreasing) had potential mean-variation relationships that could explain the association between heterogeneity and age (Table 4.2, third column). Notably, heterogeneity in physical performance measures, chronic condition count, and frailty index were not associated with age after adjustment for the mean change. Measures of 7 health characteristics had a normative or clinical scaling: chair rise and TUG times, functional measures (OARS, LSA, PASE), chronic conditions, and frailty index. Supplementary Tables 4.2-4.4 detail the intermediate results used to reach these results.

Figure 4.1 shows the association between age and effective variance for each domain, all domains, and all domains excluding laboratory values. Overall, heterogeneity increases with age. When comparing between domains, vital signs had the largest increase in heterogeneity, followed by physical performance measures and laboratory values; heterogeneity in physical measures decreases, whereas heterogeneity in functional and in physiological measures appears stable. Figures 4.1A-F show domains with significant associations between heterogeneity and age; these domains had at least second-order polynomial associations: quadratic for all domains, physical measures, laboratory values, and vital signs; cubic for all, excluding laboratory values and physical performance. The relationship was non-monotonic (i.e., varied in direction) for all domains,

laboratory values (inverted-U shape), and physical measures. Overall peak heterogeneity is reached around 70 years of age but increases continuously when excluding laboratory values (Figures 4.1A-B).

4.6.3 Comparing between-age group and within-older-age-group heterogeneity

Figure 4.2 shows the age thresholds where between-age-group deviation exceeds within-group deviation of individuals 85 years. Fifteen variables crossed this threshold between 45 and 85 years, with the thresholds for FEV1 (75.8 years), grip strength (72.5), vision (70.8), and PASE (70.5) closest to 85 years. Once this age threshold is crossed, individuals below that age can be considered more different from those 85 years old than 85-year-olds between themselves (within age group). Physical performance measures (grip strength, gait speed, chair rise, and TUG), chronic condition count, frailty index, and LSA crossed the threshold before 45 years, but not ADL-iADLs (measured by OARS).

4.7 Discussion

4.7.1 Features of heterogeneity in aging

Overall, our results confirm the widely held dictum that health heterogeneity increases with chronological age.[3,4,19] Older adults are, *in general*, more heterogeneous among themselves than younger adults. However, our analyses reveal that this statement requires many caveats. In line with, and extending previous work,[13,15,16] half of the 34 variables examined showed increased variability, but 8 showed decreased variability, and for 9 variability did not appear to change with age. Except for physical measures, heterogeneity tended to increase for all domains, but associations were mostly nonlinear, and non-monotonic for overall domains, laboratory values, and physical measures. Our findings suggest multiple heterogeneity trajectories,[3] including an inverted-U trajectory for laboratory values. Of the 17 variables with increasing heterogeneity, 8

could be attributable to mean-variation relationships, and 5 to normative or clinical scaling of measures. What is measured and how it is measured influences heterogeneity. Supplementary Table 4.5 presents 6 key features that clarify the description and understanding of heterogeneity: group, spread, measure, specificity, monotonicity, and mean-variation. Heterogeneity in aging is itself heterogeneous and multifaceted: in what follows, we wish to highlight how features of heterogeneity in older adults are relevant to clinical practice and research, as summarized in Table 4.3.

4.7.2 Clinical implications

Greater heterogeneity with age for most health characteristics and domains justifies greater attention when managing older adults.[24] The greater probability of finding clinically relevant differences in older adults compared to their younger peers strongly supports the careful and potentially time-consuming comprehensive geriatric assessment, particularly in oncological or perioperative settings where these differences are highly predictive of outcomes.[25,26]

However, this increased heterogeneity was not found for all variables and was especially important for physical performance measures, chronic condition count, frailty index, and, to a lesser extent, functional measures. These variables have in common an age-related focus and an underlying normative-clinical scaling. Assessment using age-related and clinically relevant measures will uncover greater heterogeneity in older adults. In practice, this reinforces the chief importance of physical performance, multimorbidity, frailty, and functional measures as core dimensions of the comprehensive geriatric assessment, beyond other health characteristics generally considered in the medical setting.

Our findings indicate that the scaling of measures influences the amount of heterogeneity captured within a dimension: PASE and OARS both measure function and disability, yet we show decreasing heterogeneity of PASE and increasing heterogeneity of OARS with age. PASE assesses function from extremely active to no activity whereas OARS measures ADLs which are only impaired with clinically significant functional decrease. As reduction in PASE at the higher range of functional capacity has less impact on quality of life than a reduction in OARS, the latter should be favored when evaluating older adults. Clinicians caring for older adults should select and incorporate measures that are optimally scaled for this population to better characterize heterogeneity and improve decision-making. Most age-related health characteristics are clinically scaled which may explain increasing heterogeneity with age and drive heterogeneity in health care costs.[19]

We show that increased heterogeneity in older adults can be decomposed into that between older and younger adults and between one older adult and others of the same age. Geriatric expertise and teaching are premised on this dual difference between ages and between aged individuals. These two differences contribute distinct knowledge to geriatric care: the first informs how care should be different by age groups (between the younger adults and older adults as a *group*), the second highlights the critical importance of personalizing care beyond chronological age (as an *individual* older adult in their age group). The relative importance of these two differences depends on the specific variable considered. FEV1, visual acuity, or grip strength are variables where between-age variation dominates and thus where chronological age (being "old") captures most of the variation. But for most variables the variation between older adults themselves is greater. Even in geriatric practice, chronological age can only be used as a gross surrogate marker for the mean:[16] clinicians should tailor management of older adults beyond age by

considering an individual's specific levels of prognostic factors as they will often be discordant with average age category levels. This may be contrasted with pediatric medicine where the greater part of differences is between children and adults, rather than between children themselves.

Determination of heterogeneity in aging is contingent on the referent group used. Typically, norms for vital signs and laboratory values are similar in adults regardless of age: the same referent group and center location are used from which the spread of each individual is calculated. We demonstrated that both mean value and spread about the mean changes with age for systolic blood pressure and the majority of laboratory values. This provides compelling arguments for implementation of age-specific ranges.[27–29] Moreover, individual-specific ranges to determine "normal" values could also be more widely considered, as has been recommended for temperature.[30] Using all-age or age-specific referent group may be specific to the context and depend on the expected benefit of interventions. For example, using the T-score (all-age referent group) or the Z-score (age-specific referent group) for BMD will identify individuals who may benefit differentially from treatment. For the many distribution-determined conditions, such as osteoporosis or anemia, clinicians should carefully question the appropriateness of the referent group and the clinical relevance of the absolute threshold of what is considered abnormal.

4.7.3 Research implications

Our results showing variability in increases of heterogeneity in aging preclude uniform statements about heterogeneity in older adults. They show that its quantification is intertwined with the measurement properties of instruments used (scaling and mean-variation relationship) which future research should seek to disentangle from true variation in aging. Selective survival, whereby mortality occurs in a non-random segment of a population,[33] may influence variation in heterogeneity. As extreme values of health characteristics are more strongly associated with

mortality, the attrition of individuals may result in the underestimation of heterogeneity. Homeostenosis of aging, with decreased resistance and redundancy to stressors, may translate into greater variability[34,35] but only to a threshold above which death ensues. Future research could leverage the longitudinal design of the ongoing CLSA or other cohorts to examine variability at the cohort and individual levels, and its association with mortality.

In addition to research on heterogeneity itself, our findings suggest lines of inquiry that use heterogeneity to enhance clinical management. Heterogeneity can refine the selection of variables used to develop constructs to stratify subgroups of older adults specifically. Age-related constructs, most importantly frailty, seek to capture heterogeneity among older adults as a broad group,[4] rather than distinguishing individuals among smaller subgroups of older age. We show that the heterogeneity by age decreases for many variables (e.g., FEV1, visual acuity, grip strength). To distinguish the more robust 85-year-olds from others of the same age using or developing novel scales for variables that show increasing heterogeneity by age should be considered. Conversely, measures of biological age that seek to capture the latent aging process might benefit from including variables that have *decreasing* heterogeneity by age.

Heterogeneity may inform the selection of participant subgroups for research. An epistemological and clinical assumption is that large deviation from the mean may hold potential for discovery and intervention. Modifiable health states or trajectories are more likely to be identified in individuals with outlying characteristics *from their age group* rather than outlying from all adults, especially since adults of considerable age may all be outliers from the general population.

Overall heterogeneity appears to have an inverted U-shape with maximum variability at approximately 70 years. This inverted shape is strongly driven by laboratory values, raising the

possibility that standard laboratory measures optimally distinguish younger older adults and that other better suited biomarkers should be developed for older age groups.

4.7.4 Limitations

First, due to the large number of variables, our exploratory results may be prone to multiple testing issues. Nonetheless, most reported associations had strong statistical significance and were also clinically significant. Second, participants were community-dwelling older adults without cognitive impairment. Our analyses may underestimate heterogeneity if institutionalized and/or cognitively impaired older adults have more extreme variable values. However, the age range of participants from 45 to 85 years allowed exploration of heterogeneity in younger age groups where the proportion of excluded participants was low. Third, although we attempted to choose variables representative of clinical practice, our selection of health characteristics may have influenced our findings. We focused on health states, but heterogeneity has also been described on psychological and social levels.[15,16] Fourth, because our analyses were cross-sectional, we cannot disentangle period or cohort effects from the true aging process *per se*.[3] Our findings should not be considered from the perspective of mechanistic or biological aging, but from a perspective of descriptive aging, which holds a predictive and clinically relevant meaning as discussed above. Along the same lines, we did not account for clinical management which may decrease "natural" variability for some variables (e.g., HbA1C, TSH, LDL). From a descriptive standpoint, treatment can be understood as a valid modifier of observed variation in heterogeneity with *chronological* age: medical conditions are increasingly prevalent with age but are also treated.

4.8 Conclusion

Overall health heterogeneity increases with age but does not uniformly increase across all variables and domains. Heterogeneity in aging reinforces the need for geriatric assessment and care, depends

on which health characteristics are assessed, their measurement properties, and their referent group. Like the older adults it seeks to describe, heterogeneity in aging is itself heterogeneous, suggesting further research to develop improved single-dimension and multidimensional instruments, as well as specific vital and laboratory reference ranges for older adults.

4.9 ACKNOWLEDGEMENTS

The authors wish to thank Drs. Mark Clarfield and Han Ting Wang for their suggestions to improve the manuscript. This research was made possible using the data/biospecimens collected by the Canadian Longitudinal Study on Aging (CLSA). Funding for the Canadian Longitudinal Study on Aging (CLSA) is provided by the Government of Canada through the Canadian Institutes of Health Research (CIHR) under grant reference: LSA 94473 and the Canada Foundation for Innovation. This research has been conducted using the CLSA dataset Comprehensive V4.0, under Application Number 190209. The CLSA is led by Drs. Parminder Raina, Christina Wolfson and Susan Kirkland. The opinions expressed in this manuscript are the author's own and do not reflect the views of the Canadian Longitudinal Study on Aging.

Sponsor's Role

The sponsor had no role in the design and conduct of this study; analysis and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Author contributions

QDN, EMM, CW designed the study. QDN analyzed data and drafted the manuscript. All authors contributed significantly to the content, critically reviewed, and approved the final manuscript for publication.

Conflict of interests

None.

4.10 REFERENCES

1. Freedman VA, Spillman BC. Disability and care needs among older Americans. Milbank Q. 2014;92(3):509–541. doi:10.1111/1468-0009.12076

2. Vaupel JW, Manton KG, Stallard E. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*. 1979;16(3):439. doi:10.2307/2061224

3. Dannefer D, Sell RR. Age structure, the life course and "aged heterogeneity": prospects for research and theory. *Compr Gerontol B*. 1988;2(1):1–10.

4. Mitnitski A, Howlett SE, Rockwood K. Heterogeneity of human aging and its assessment. *J Gerontol A Biol Sci Med Sci*. 2016;72(7):877–884. doi:10.1093/gerona/glw089

5. Oxford English Dictionary. "Heterogeneity." https://www.lexico.com/definition/heterogeneity.

6. Carnes BA, Olshansky SJ. Heterogeneity and its biodemographic implications for longevity and mortality. *Exp Gerontol*. 2001;36(3):419–430. doi:10.1016/S0531-5565(00)00254-0

7. Mitnitski AB, Mogilner AJ, Rockwood K. Accumulation of deficits as a proxy measure of aging. *Sci World*. 2001;1:323-336. doi:10.1100/tsw.2001.58

8. Fried LP, Tangen CM, Walston J, et al. Frailty in older adults: evidence for a phenotype. *J Gerontol A Biol Sci Med Sci*. 2001;56(3):808–813. doi:10.1093/gerona/56.3.M146

9. Fabbri E, Zoli M, Gonzalez-Freire M, Salive ME, Studenski SA, Ferrucci L. Aging and multimorbidity: new tasks, priorities, and frontiers for integrated gerontological and clinical research. *J Am Med Dir Assoc*. 2015;16(8):640–647. doi:10.1016/j.jamda.2015.03.013

10. Nguyen QD, Wu C, Odden MC, Kim DH. Multimorbidity patterns, frailty, and survival in community-dwelling older adults. *J Gerontol A Biol Sci Med Sci*. 2019;74(8):1265–1270. doi:10.1093/gerona/gly205/5088142

11. Hoogendijk EO, Afilalo J, Ensrud KE, Kowal P, Onder G, Fried LP. Series Frailty 1 Frailty : implications for clinical practice and public health. *Lancet*. 2019;394(10206):1365–1375. doi:10.1016/S0140-6736(19)31786-6

12. Botwinick J, Thompson LW. A research note on individual differences in reaction time in relation to age. *J Genet Psychol*. 1968;112:73-75.

13. Maddox GL, Douglass EB. Aging and individual differences: a longitudinal analysis of social, psychological, and physiological indicators. *J Gerontol*. 1974;29(5):555–563. doi:10.1093/geronj/29.5.555

14. Bornstein R, Smircina MT. The status of the empirical support for the hypothesis of increased variability in aging populations. *Gerontologist*. 1982;22(3):258–260. doi:10.1093/geront/22.3.258

15. Nelson AE, Dannefer D. Aged heterogeneity: fact or fiction? The fate of diversity in gerontological research. *Gerontologist*. 1992;32(1):17–23. doi:10.1093/geront/32.1.17

16. Stone ME, Lin J, Dannefer D, Kelley-Moore JA. The continued eclipse of heterogeneity in gerontological research. *J Gerontol B Psychol Sci Soc Sci*. 2017;72(1):162-167. doi:10.1093/geronb/gbv068

17. Santoni G, Angleman S, Welmer AK, Mangialasche F, Marengoni A, Fratiglioni L. Age-related variation in health status after age 60. *PLoS One*. 2015;10(3):1–10. doi:10.1371/journal.pone.0120077

18. Mungas D, Beckett L, Harvey D, et al. Heterogeneity of cognitive trajectories in diverse older persons. *Psychol Aging*. 2010;25(3):606–619. doi:10.1037/a0019502

19. Lowsky DJ, Olshansky SJ, Bhattacharya J, Goldman DP. Heterogeneity in healthy aging. *J Gerontol A Biol Sci Med Sci*. 2014;69(6):640–649. doi:10.1093/gerona/glt162

20. Raina P, Wolfson C, Kirkland S, et al. Cohort Profile: The Canadian Longitudinal Study on Aging (CLSA). *Int J Epidemiol*. 2019;48(6):1752–1753j. doi:10.1093/ije/dyz173

21. Raina PS, Wolfson C, Kirkland SA. Canadian Longitudinal Study on Aging - Protocol. https://clsa-elcv.ca/doc/511. Accessed November 26, 2019.

22. Searle SD, Mitnitski A, Gahbauer EA, Gill TM, Rockwood K. A standard procedure for creating a frailty index. *BMC Geriatr*. 2008;8(1):24. doi:10.1186/1471-2318-8-24

23. Peña D, Rodríguez J. Descriptive measures of multivariate scatter and linear dependence. *J*

Multivar Anal. 2003;85(2):361–374. doi:10.1016/S0047-259X(02)00061-1

24. Parker SG, McCue P, Phelps K, et al. What is Comprehensive Geriatric Assessment (CGA)? An umbrella review. *Age Ageing*. 2018;47(1):149–155. doi:10.1093/ageing/afx166

25. Chow WB, Rosenthal RA, Merkow RP, Ko CY, Esnaola NF. Optimal preoperative assessment of the geriatric surgical patient: A best practices guideline from the American college of surgeons national surgical quality improvement program and the American geriatrics society. *J Am Coll Surg*. 2012;215(4):453–466. doi:10.1016/j.jamcollsurg.2012.06.017

26. Wildiers H, Heeren P, Puts M, et al. International society of geriatric oncology consensus on geriatric assessment in older patients with cancer. *J Clin Oncol*. 2014;32(24):2595–2603. doi:10.1200/JCO.2013.54.8347

27. Chester JG, Rudolph JL. Vital signs in older patients: age-related changes. *J Am Med Dir Assoc*. 2011;12(5):337–343. doi:10.1016/j.jamda.2010.04.009

28. Vadiveloo T, Donnan PT, Murphy MJ, Leese GP. Age- and gender-specific TSH reference intervals in people with no obvious thyroid disease in Tayside, Scotland: The thyroid epidemiology, audit, and research study (TEARS). *J Clin Endocrinol Metab*. 2013;98(3):1147-1153. doi:10.1210/jc.2012-3191

29. Santoni G, Calderón-Larrañaga A, Vetrano DL, Welmer A-K, Orsini N, Fratiglioni L. Geriatric Health Charts for Individual Assessment and Prediction of Care Needs: A Population-Based Prospective Study. *J Gerontol A Biol Sci Med Sci*. 2020;75(1):131–138. doi:10.1093/gerona/gly272

30. High KP, Bradley SF, Gravenstein S, et al. Clinical practice guideline for the evaluation of fever and infection in older adult residents of long-term care facilities: 2008 Update by the Infectious Diseases Society of America. *Clin Infect Dis*. 2009;48(2):149–171. doi:10.1086/595683

31. Williams DM, Jylhävä J, Pedersen NL, Hägg S. A frailty index for UK Biobank participants. *J Gerontol A Biol Sci Med Sci*. 2019;74(4):582–587. doi:10.1093/gerona/gly094

32. Sourial N, Bergman H, Karunananthan S, et al. Implementing frailty into clinical practice:

a cautionary tale. *J Gerontol A Biol Sci Med Sci*. 2013;68(12 A):1505–1511. doi:10.1093/gerona/glt053

33. Markides KS, Machalek R. Selective survival, aging and society. *Arch Gerontol Geriatr*. 1984;3(3):207–222. doi:10.1016/0167-4943(84)90022-0

34. Ferrucci L, Windham BG, Fried LP. Frailty in older persons. *Genus*. 2005;61(1):39–53.

35. Lipsitz LA, Goldberger AL. Loss of “Complexity” and Aging: Potential Applications of Fractals and Chaos Theory to Senescence. *JAMA*. 1992;267(13):1806–1809. doi:10.1001/jama.1992.03480130122036

4.11 TABLES

Table 4.1. Characteristics of community-dwelling older adults in the Canadian Longitudinal Study on Aging Comprehensive cohort, by age group

	Overall	45–54Years	55–64 Years	65–74 Years	75–86 Years
Sample size, n	30,097	7595	9856	7362	5284
Age, years,	63.0 (10.2)	50.3 (2.7)	59.7 (2.8)	68.9 (2.8)	78.9 (2.9)
Sex, male, n	14777 (49.1)	3670 (48.3)	4767 (48.4)	3674 (49.9)	2666 (50.5)
Race, White, n	28771 (95.6)	7098 (93.5)	9463 (96.0)	7097 (96.4)	5113 (96.8)
Physical measures,					
Weight, kg	79.7 (17.6)	80.9 (18.8)	81.3 (18.3)	79.6 (16.7)	75.4 (14.8)
Height, m	1.68 (0.10)	1.70 (0.09)	1.69 (0.10)	1.67 (0.10)	1.66 (0.10)
BMI, kg/m²	28.1 (5.4)	27.8 (5.7)	28.4 (5.7)	28.3 (5.3)	27.3 (4.6)
Waist circumference, cm	94 (15)	92 (15)	95 (15)	96 (14)	95 (13)
Vital signs					
Pulse, beats per min	72 (12)	73 (11)	72 (12)	71 (12)	70 (12)
Systolic blood pressure, mmHg	122 (17)	116 (15)	121 (16)	126 (17)	128.6 (18)
Diastolic blood pressure, mmHg	74 (10)	76 (10)	76 (10)	73.6 (10)	71 (10)
Physiological measures					
FEV1, L	2.7 (0.8)	3.1 (0.7)	2.8 (0.7)	2.5 (0.7)	2.1 (0.6)
BMD, g/cm²	1.01 (0.14)	1.03 (0.12)	1.00 (0.13)	1.00 (0.15)	0.99 (0.16)
Visual acuity*	0.9 (0.3)	1.0 (0.3)	0.9 (0.3)	0.8 (0.3)	0.7 (0.2)
Physical performance measures					
Gait speed, m/s	0.98 (0.20)	1.04 (0.18)	1.01 (0.20)	0.95 (0.19)	0.86 (0.19)
Grip strength, kg	35.2 (11.8)	39.4 (12.3)	36.2 (11.6)	33.4 (10.8)	29.0 (9.8)
Chair rise, s for one†	2.7 (0.8)	2.5 (0.7)	2.6 (0.7)	2.8 (0.8)	3.0 (0.9)
TUG, s	9.6 (2.6)	8.8 (1.7)	9.2 (2.4)	9.8 (2.3)	11.2 (3.4)
Function and disability					
ADL impairment, n‡	837 (2.8)	95 (1.3)	210 (2.1)	198 (2.7)	334 (6.3)
IADL impairment, n‡	1566 (5.2)	170 (2.2)	345 (3.5)	371 (5.1)	680 (12.9)
OARS score§	27.8 (0.7)	27.9 (0.6)	27.9 (0.6)	27.8 (0.7)	27.6 (0.9)
Life Space Assessment score	85 (18)	91 (17)	87 (18)	83 (18)	77 (19)
PASE score	141 (74)	174 (83)	149 (73)	125 (60)	100 (53)
Chronic conditions, n					
Hypertension	11101 (37.0)	1495 (19.7)	3364 (34.2)	3369 (45.9)	2873 (54.6)
Diabetes	2957 (9.9)	370 (4.9)	963 (9.9)	923 (12.7)	701 (13.6)
Heart disease	4232 (14.1)	317 (4.2)	1003 (10.2)	1383 (18.8)	1529 (29.1)
Stroke or TIA	1347 (4.5)	101 (1.3)	282 (2.9)	395 (5.4)	569 (10.8)

Arthritis	7922 (26.4)	926 (12.2)	2472 (25.1)	2514 (34.2)	2010 (38.2)
Osteoporosis	2689 (9.0)	154 (2.0)	738 (7.5)	931 (12.7)	866 (16.5)
Lung disease	5094 (17.0)	1265 (16.7)	1687 (17.1)	1268 (17.3)	874 (16.6)
Kidney disease	867 (2.9)	99 (1.3)	244 (2.5)	252 (3.4)	272 (5.2)
Cancer	4637 (15.4)	427 (5.6)	1270 (12.9)	1454 (19.8)	1486 (28.2)
Anxiety or depression	6243 (20.8)	1736 (22.9)	2356 (23.9)	1464 (19.9)	687 (13.1)
Chronic condition count, mean	1.6 (1.4)	0.9 (1.0)	1.5 (1.3)	1.9 (1.4)	2.3 (1.5)
Frailty index	0.09 (0.08)	0.05 (0.06)	0.08 (0.07)	0.10 (0.08)	0.14 (0.09)
Laboratory values					
Hemoglobin, g/L	141 (13)	141 (13)	142 (13)	141 (13)	138 (14)
WBC count, 10^9/L	6.7 (2.2)	6.6 (1.9)	6.6 (1.8)	6.8 (2.3)	7.0 (3.1)
Platelet count, 10^9/L	222 (58)	227 (56)	225 (58)	220 (59)	211 (58)
Creatinine, μmol/L	82 (24)	79 (20)	80 (25)	83 (23)	88 (27)
GFR, mL/s/m²	79 (15)	88 (13)	82 (13)	75 (14)	66 (14)
ALT, U/L	24 (14)	25 (14)	25 (15)	23 (13)	20 (12)
Albumin, g/L	40 (3)	40 (3)	40 (3)	40 (3)	39 (3)
Total cholesterol, mmol/L	5.1 (1.1)	5.2 (1.0)	5.3 (1.1)	5.0 (1.2)	4.8 (1.2)
HDL, mmol/L	1.5 (0.5)	1.5 (0.5)	1.5 (0.5)	1.5 (0.5)	1.5 (0.5)
LDL, mmol/L	2.8 (1.0)	3.0 (0.9)	3.0 (1.0)	2.8 (1.0)	2.6 (1.0)
TSH, mU/L	2.3 (2.2)	2.2 (1.8)	2.2 (2.6)	2.3 (1.7)	2.5 (2.4)
HbA1C, %	5.7 (0.8)	5.5 (0.7)	5.7 (0.8)	5.7 (0.7)	5.8 (0.7)
Ferritin, μg/L	158 (143)	142 (141)	162 (143)	168 (150)	161 (136)
Vitamin D, nmol/L	90 (38)	79 (34)	88 (37)	96 (39)	99 (38)
C-reactive protein, mg/L	2.6 (5.1)	2.2 (4.1)	2.5 (4.9)	2.7 (5.8)	2.9 (5.5)

Notes: (%) or (SD) as appropriate; ***** Fraction (e.g., 20/20 = 1); † Average time required for one chaise rise; ‡ Impairment on one or more ADL or iADL; § Older Americans Resources and Services score of 14 ADL and iADL (0–2) ;ALT = alanine aminotransferase; BMI = body mass index; FEV1 = forced expiratory volume in 1 second; BMD = bone mineral density; ADL = activity of daily living; GFR = glomerular filtration rate; HDL = high-density lipoprotein; IADL = instrumental activity of daily living; LDL = low-density lipoprotein; OARS = Older Americans Resources and Services subscales for 7 activities of daily living (ADL) and 7 instrumental ADLs; PASE = Physical activity scale for the elderly; TUG = Timed up and go; TSH = thyroid-stimulating hormone; WBC = white blood cell

Table 4.2. Variation in heterogeneity and mean by health characteristic, in relation to chronological age

Health characteristic	Variation in Heterogeneity by Age[†] (Deviation Slope, from Model 1)	Variation in Mean by Age[†] (Mean Slope, from Model 2)	Variation Explainable by Mean-Variation Relationship[‡] (from Model 3)	Measured Characteristic Has Normative or Clinical Scaling
Physical measures				
Weight	-	-	No	No
Height	↔	-		No
BMI	-	-	No	No
Waist circumference	-	+	No	No
Vital signs				
Pulse	↔	-		No
Systolic blood pressure	+	+	No	No
Diastolic blood pressure	↔	-		No
Physiological measures				
FEV1	-	-	No	No
BMD	+	-	No	No
Visual acuity	-	-	Yes	No
Physical performance measures				
Gait speed	↔	-		No
Grip strength	-	-	Yes	No
Chair rise	+	+	Yes	Yes
TUG	+	+	Yes	Yes
Function and disability				
OARS	+	-	No	Yes
Life Space Assessment	+	-	No	Yes
PASE	-	-	Yes	Yes
Chronic condition count	+	+	Yes	Yes
Frailty index	+	+	Yes	Yes
Laboratory values				
Hemoglobin	+	-	No	No
WBC count	↔	↔		No
Platelet count	↔	-		No
Creatinine	+	+	No	No
GFR	+	-	No	No
ALT	-	-	No	No
Albumin	↔	-		No
Total cholesterol	+	-	No	No
HDL	↔	↔		No
LDL	+	-	No	No
TSH	+	+	Yes	No
HbA1C	+	+	Yes	No
Ferritin	+	+	Yes	No
Vitamin D	+	+	Yes	No
C-reactive protein	↔	+		No

Notes. "+" indicates an increase, "-" decrease, "↔" not clinically significant. † Variation determined by linear regression of deviation (model 1) or mean (model 2) on age, adjusted for sex (sex-specific intercepts). ‡ Determined by examining the mean-deviation relationship and adjusting linear regressions for age and sex group mean; see Methods and Supplementary Figure 4.1 for details.

Table 4.3. Summary of major findings on heterogeneity in aging, clinical and research implications

Findings	Clinical implications	Research implications and themes
Overall heterogeneity in health increases with chronological age.	- Heterogeneity underlies the need and relevance of age-appropriate care and management.	-
Heterogeneity does not increase uniformly and may be attributable measurement properties (e.g., scaling of measures, mean-variation relationship).	-	- Variability in heterogeneity in aging precludes uniform statement about heterogeneity in older adults. - Investigate the impact of measurement properties and selective survival on variation in heterogeneity in aging.
Clinically age-relevant variables are more heterogeneous with age.	- Heterogeneity reinforces the importance of clinically age-relevant variables in the CGA (chronic conditions, function and disability, physical performance measures, frailty, etc.).	- Develop heterogeneity capturing measures that use optimal scaling for older adults.
The scaling of a measure determines to a great extent the amount of heterogeneity detected.	- Clinicians should select measures which use clinically relevant scaling for its intended purpose, e.g., PASE vs. OARS. - Heterogeneity of health care costs with age is driven by clinically relevant measures.	
Heterogeneity in aging can be decomposed into differences between age groups and differences between older individuals (within age group).	- Geriatric care is based on managing older adults differently from younger adults as well as differently between older adults themselves. - Care for older adults must account for age while also going beyond age as a surrogate mean marker for relevant prognostic factors.	- Refine and develop multidimensional constructs to stratify subgroups older adults by using variables that are most heterogeneous among them. - Measures of biological age may benefit from using variables that are less heterogeneous with aging.
Deviation from the mean (and heterogeneity) will vary by the specific group of reference selected. This is especially true when there are important differences between the mean values by age group.	- The reference group selected is essential to interpret disease states and conditions that are based on a statistical distribution, e.g., osteoporosis (T-score vs. Z-score), vital signs, or anemia. - Underdiagnosis or overdiagnosis may occur if a younger referent group is used without relevant clinical justification.	- Identify participants for research as those outlying within their chronological age group. - Investigate alternative and clinically useful ranges for vital signs and laboratory results for older adults.
Laboratory values attain peak heterogeneity in the late sixties.	-	- Develop and integrate laboratory biomarkers that are better suited to assess and differentiate older adults.

4.12 FIGURES

Figure 4.1. Nonlinear variation in effective variance by chronological age, overall and by health domains

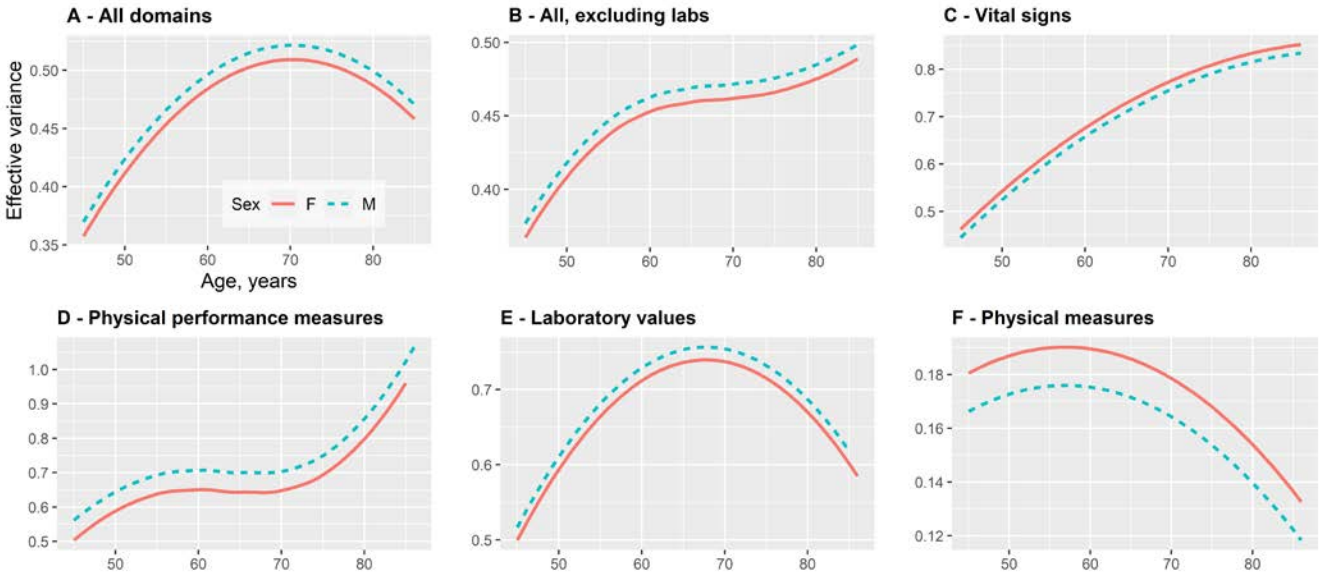


Figure 4.1 Legend

Predicted effective variance curves overall and by domains, with sex-specific intercepts, are illustrated for significant (nonlinear) associations. The associations between chronological age and effective variance by Function and disability measures and by Physiological measures were non-significant.

Figures 4.1A-F. Figures 4.1A and 4.1E. Predicted effective variance curves reveal clearly non-monotonic relationships where heterogeneity increases until approximately 70 years and then decreases for All domains and Laboratory values. Although age is linearly associated with heterogeneity for both, assuming a linear relationship is misleading due to non-monotonicity. Figures 4.1B, 4.1C, and 4.1D. Heterogeneity increases with age for All, excluding laboratory values, Vital signs, and Physical performance measures. Only the effective variance of Vital signs increases approximately linearly. Effective variance for Physical performance measures appears to stabilize between 55 and 70 years old.

Figure 4.1F. Effective variance for Physical measures appears to peak around 57 years and then decrease.

Figure 4.2. Age thresholds at which mean deviation between an age group and the 85-year-old age group exceeds mean deviation within the group of 85-year-old individuals

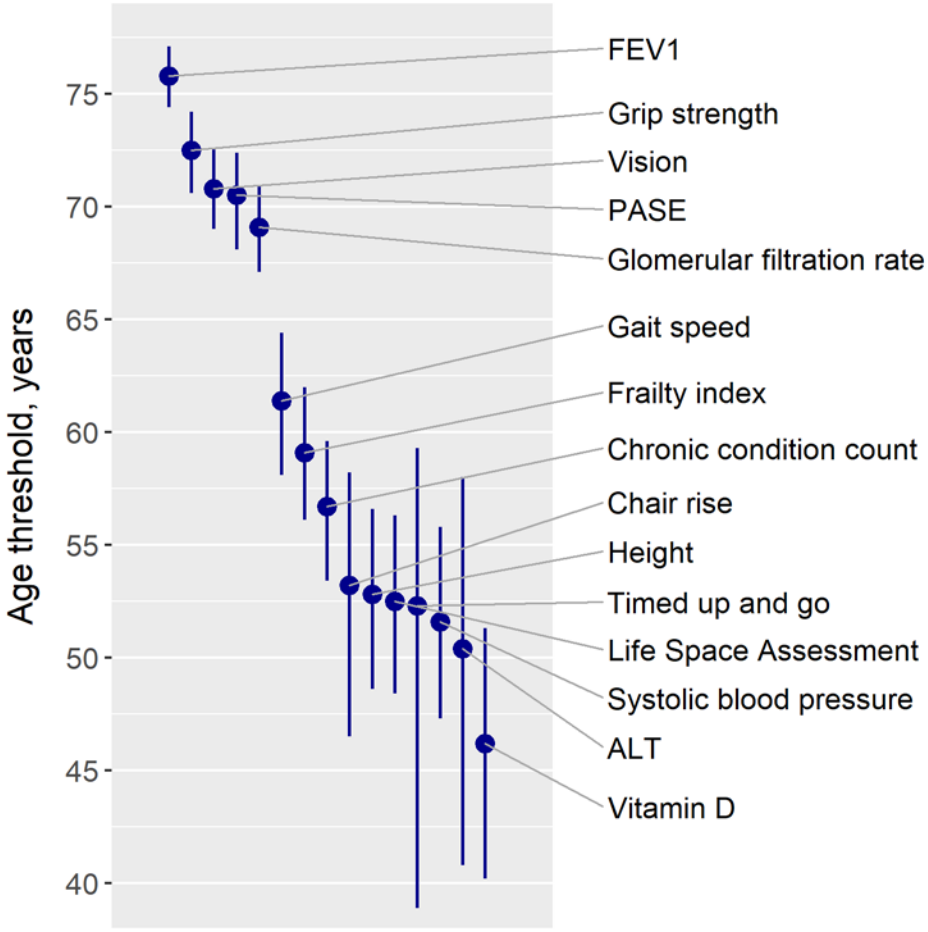


Figure 4.2 Legend

FEV1 = forced expiratory volume in 1 second; PASE = Physical activity scale for the elderly. Vertical lines indicate bootstrapped 95% confidence intervals.

The age thresholds give an indication of the relationship between variation *within* the 85-year-old age group compared to variation *between* age groups for each health characteristic. Once this age threshold is crossed, individuals below that age can be considered more different from those 85 years old than 85-year-olds between themselves.

For example for *FEV1*, starting at 76 years, the deviation between the average 85-year-old individual and the average 76-year-old individual ("between-group" deviation) exceeds the mean deviation between 85-year-old individuals themselves ("within-group" deviation). The analogous threshold is crossed at 46 years for *Vitamin D*. Fifteen health characteristics cross the threshold between 45 and 85 years. The remaining health characteristics do not: *albumin, BMI, bone mineral density, cholesterol, CRP, creatinine, diastolic blood pressure, ferritin, HbA1C, HDL,*

hemoglobin, LDL, OARS28, platelet, pulse, TSH, waist circumference, weight, WBC. For these health characteristics, individuals at 85 years continue to show more deviation within their age group than between their age group and the age group of individuals aged 45 years.

For most variables, the variation between older adults is greater than between age groups: this requires clinicians to tailor management based on specific individual values of prognostic health characteristics rather than relying on chronological age.

See Methods and Supplementary Figure 4.2 for more details.

S4.13 Supplemental Material

Supplementary Methods 4.1

Health characteristics examined for heterogeneity in aging

Domains	Variables
Physical measures (4 variables)	Weight Height BMI Waist circumference
Vital signs (3)	Pulse Systolic blood pressure Diastolic blood pressure
Physiological measures (3)	Forced expiratory volume in 1 second (FEV1), Bone mineral density (BMD) Vision impairment
Physical performance measures (4)	Gait speed Grip strength Chair rise Timed up and go (TUG)
Function and disability (3)	Older Americans Resources and Services (OARS) subscales for 7 activities of daily living (ADL) and 7 instrumental ADLs (iADL) (range = 0–28) Life space assessment (LSA) Physical activity scale for the elderly (PASE)
Chronic condition count (1)	Count of 10 self-reported chronic conditions: hypertension, diabetes, heart disease (heart disease, angina, coronary bypass, or angioplasty), stroke or transient ischemic attack, arthritis, osteoporosis, lung disease, kidney disease, cancer, anxiety/depression (range = 0-10)
Frailty index (1)	Count of: [OARS subscales (14 counts), 10 comorbidities (above), low activity, exhaustion, grip strength, gait speed, weight loss, health perception, fall, FEV1, self-reported hearing impairment, self-reported vision impairment] ÷ 34
Laboratory values (15)	Hemoglobin level, white blood cell count (WBC), platelet count, alanine aminotransferase (ALT), albumin, creatinine, glomerular filtration rate (GFR), total cholesterol, HDL, LDL, thyroid stimulating hormone (TSH), hemoglobin A1C (HbA1C), ferritin, hydroxyvitamin D, high sensitivity C-reactive protein (CRP)

Variables used to define the frailty index

OARS subscales (14 counts), 10 comorbidities (above), low activity, exhaustion, grip strength, gait speed, weight loss, health perception, fall, FEV1, self-reported hearing impairment, self-reported vision impairment

Supplementary Table 4.1. Missing data and proportion for health characteristics

Health characteristics	Count missing (n)	Proportion missing (n = 30,097)
FEV1	7276	0.242
Hemoglobin	4670	0.155
WBC	4670	0.155
Platelet	4670	0.155
LDL	3671	0.122
HbA1C	3187	0.106
Vitamin D	3092	0.103
ALT	3091	0.103
Ferritin	3086	0.103
Creatinine	3085	0.103
Glomerular filtration rate	3085	0.103
Albumin	3085	0.103
Cholesterol	3085	0.103
HDL	3085	0.103
TSH	3085	0.103
CRP	3085	0.103
Grip strength	2290	0.076
PASE	1633	0.054
Chair rise	1334	0.044
Bone mineral density	1317	0.044
Timed up and go	430	0.014
Gait speed	392	0.013
Chronic condition count	349	0.012
Vision	304	0.010
Pulse	289	0.010
Systolic blood pressure	286	0.010
Diastolic blood pressure	286	0.010
Waist circumference	235	0.008
OARS28	160	0.005
BMI	136	0.005
Weight	130	0.004
Height	100	0.003
Frailty index	64	0.002
Life Space Assessment	53	0.002

Notes. ALT = alanine aminotransferase; BMI = body mass index; FEV1 = forced expiratory volume in 1 second; BMD = bone mineral density; ADL = activity of daily living; GFR = glomerular filtration rate; HDL = high-density lipoprotein; IADL = instrumental activity of daily living; LDL = low-density lipoprotein; OARS = Older Americans Resources and Services; PASE = Physical activity scale for the elderly; TUG = Timed up and go; TSH = thyroid-stimulating hormone; WBC = white blood cell

Supplementary Table 4.2. Coefficients for change in deviation by age, statistical and clinical significance

	Model 1 Statistical Significance				Clinical Significance		
Health Characteristic	**Coefficient for Deviation per 10y**	**p**	**Coefficient for Male**	**p**	**Last Clinically Relevant Unit (LCRU) Criteria**	**Clinical Significance: Change in Deviation Over 40y > LCRU**	**Combined Significance**
Physical measures							
Weight	-0.932	4.52E-58	-0.293	0.0136	1	Yes	-
Height	-0.000505	0.0266	0.00513	3.85E-28	0.01	No	↔
BMI	-0.248	5.23E-36	-1.03	1.37E-140	0.1	Yes	-
Waist circumference	-0.434	1.89E-20	-1.22	4.34E-37	1	Yes	-
Vital signs							
Pulse	0.0188	0.638	0.95	4.12E-31	1	No	↔
Systolic blood pressure	1.16	7.06E-90	-0.408	0.000573	1	Yes	+
Diastolic blood pressure	0.0294	0.371	0.131	0.0525	1	No	↔
Physical measures							
Gait speed	0.00264	0.000148	-0.00413	0.00379	0.1	No	↔
Grip strength	-0.338	5.75E-39	2.68	0	1	Yes	-
Chair rise	0.0444	5.14E-46	-0.0246	0.0000981	0.1	Yes	+
TUG	0.302	2.98E-171	-0.0349	0.114	1	Yes	+
Physiological measures							
FEV1	-0.0153	3E-13	0.145	4.8E-255	0.01	Yes	-
BMD	0.00909	1.81E-109	0.0117	4.25E-45	0.01	Yes	+
Visual acuity	-0.0129	1.75E-40	0.0221	1.33E-28	0.01	Yes	-
Function and disability							
OARS	0.125	5.91E-271	-0.194	1.8E-158	0.1	Yes	+
Life Space Assessment	0.498	1.67E-16	-0.639	2.47E-07	1	Yes	+
PASE	-9.81	0	6.61	9.25E-42	1	Yes	-
Chronic condition count	0.124	1.49E-171	-0.0624	5.15E-12	0.1	Yes	+
Frailty index	0.00768	9.89E-186	-0.00909	6.73E-64	0.01	Yes	+
Laboratory values							
Hemoglobin	0.633	1.11E-44	1.04	2.14E-29	1	Yes	+
WBC count	0.0273	0.0117	0.0164	0.458	1	No	↔
Platelet count	0.221	0.316	-4.57	3.35E-24	1	Yes	↔
Creatinine	2.11	8.56E-84	2.9	4.37E-39	1	Yes	+
GFR	0.402	1.58E-18	-0.247	0.0081	1	Yes	+
ALT	-0.842	8.78E-39	1.83	1.36E-43	1	Yes	-
Albumin	0.0131	0.18	-0.0808	0.0000544	1	No	↔

	Model 1 Statistical Significance				Clinical Significance		
Health Characteristic	**Coefficient for Deviation per 10y**	**p**	**Coefficient for Male**	**p**	**Last Clinically Relevant Unit (LCRU) Criteria**	**Clinical Significance: Change in Deviation Over 40y > LCRU**	**Combined Significance**
Total cholesterol	0.0369	2.14E-22	0.0201	0.00924	0.1	Yes	+
HDL	0.00559	0.000671	-0.0726	4.52E-103	0.1	No	↔
LDL	0.0285	2.21E-17	0.000685	0.92	0.1	Yes	+
TSH	0.062	7.87E-08	-0.0609	0.00982	0.1	Yes	+
HbA1C	0.034	1.91E-20	0.115	7.56E-53	0.1	Yes	+
Ferritin	3.84	3.65E-10	47.7	0	1	Yes	+
Vitamin D	1.49	2.13E-26	-2.24	4.43E-15	1	Yes	+
C-reactive protein	0.186	6.83E-12	-0.368	3.18E-11	1	No	↔

Supplementary Table 4.3. Coefficients for change in mean by age, statistical and clinical significance

Health Characteristic	Model 2 Statistical Significance: Coefficient for Mean per 10y	p	Clinical Significance: Change in Mean Over 40y > LCRU	Combined Significance
Physical measures				
Weight	-1.92	7.55E-100	Yes	-
Height	-0.0164	0	Yes	-
BMI	-0.128	0.0000313	Yes	-
Waist circumference	1.12	5.66E-50	Yes	+
Vital signs				
Pulse	-1.08	1.87E-61	Yes	-
Systolic blood pressure	4.42	0	Yes	+
Diastolic blood pressure	-1.86	4.03E-253	Yes	-
Physical performance measures				
Gait speed	-0.0627	0	Yes	-
Grip strength	-3.71	0	Yes	-
Chair rise	0.186	0	Yes	+
TUG	0.815	0	Yes	+
Physiological measures				
FEV1	-0.357	0	Yes	-
BMD	-0.0156	4.95E-131	Yes	-
Visual acuity	-0.116	0	Yes	-
Function and disability				
OARS	-0.0902	7.77E-113	Yes	-
Life Space Assessment	-4.73	0	Yes	-
PASE	-25.9	0	Yes	-
Chronic condition count	0.472	0	Yes	+
Frailty index	0.0285	0	Yes	+
Laboratory values				
Hemoglobin	-1.22	5.08E-66	Yes	-
WBC count	0.152	2.37E-28	No	↔
Platelet count	-5.11	3.99E-50	Yes	-
Creatinine	2.91	1.2E-110	Yes	+
GFR	-7.64	0	Yes	-
ALT	-1.53	6.82E-80	Yes	-
Albumin	-0.412	1.08E-146	Yes	-
Total cholesterol	-0.143	7.85E-111	Yes	-
HDL	0.0168	1.45E-10	No	↔
LDL	-0.145	2.63E-143	Yes	-
TSH	0.106	2.77E-16	Yes	+
HbA1C	0.115	2.18E-143	Yes	+
Ferritin	6.46	2.23E-15	Yes	+
Vitamin D	7.21	4.04E-235	Yes	+
C-reactive protein	0.242	1.17E-15	Yes	+

Supplementary Table 4.4. Explainability of change in deviation by mean-variability relation

			Model 3							
Health Characteristic	**Combined Significance?**	**Same Sign Coefficient Models 1 and 2?**	**Coefficient Dev. Per 10y**	**p**	**Coefficient Dev. By Mean (by age, sex)**	**p**	**Coefficient Dev. By Sex**	**p**	**Different sign or loss of significance after adjustment**	**Variation Explainable by Mean-Variability Relation**
Physical measures										
Weight	Yes	Yes	-0.377	0	0.27	0	-4.186	0	No	No
Height	No									
BMI	Yes	Yes	-0.197	0	0.313	0	-1.164	0	No	No
Waist circumference	Yes	No								No
Vital signs										
Pulse	No									
Systolic blood pressure	Yes	Yes	1.044	0	0.02	0.492	-0.454	0.001	No	No
Diastolic blood pressure	No									
Physical performance measures										
Gait speed	No									
Grip strength	Yes	Yes	0.08	0.426	0.113	0	0.715	0.119	Yes	Yes
Chair rise	Yes	Yes	-0.025	0.006	0.349	0	-0.002	0.786	Yes	Yes
TUG	Yes	Yes	0.026	0.194	0.296	0	-0.037	0.012	Yes	Yes
Physiological measures										
FEV1	Yes	Yes	-0.035	0.023	-0.055	0.199	0.195	0	No	No
BMD	Yes	No								No
Visual acuity	Yes	Yes	0.019	0.015	0.272	0	0.01	0.003	Yes	Yes
Function and disability										
OARS	Yes	No								No
Life Space Assessment	Yes	No								No
PASE	Yes	Yes	-1.838	0.243	0.304	0	0.603	0.65	Yes	Yes
Chronic condition count	Yes	Yes	-0.041	0.058	0.349	0	0.026	0.067	Yes	Yes
Frailty index	Yes	Yes	0.001	0.179	0.214	0	-0.004	0	Yes	Yes
Laboratory values										
Hemoglobin	Yes	No								No
WBC count	No									
Platelet count	No									
Creatinine	Yes	Yes	0.604	0	0.478	0	-6.507	0	No	No
GFR	Yes	No								No
ALT	Yes	Yes	-0.26	0	0.35	0	0.019	0.912	No	No

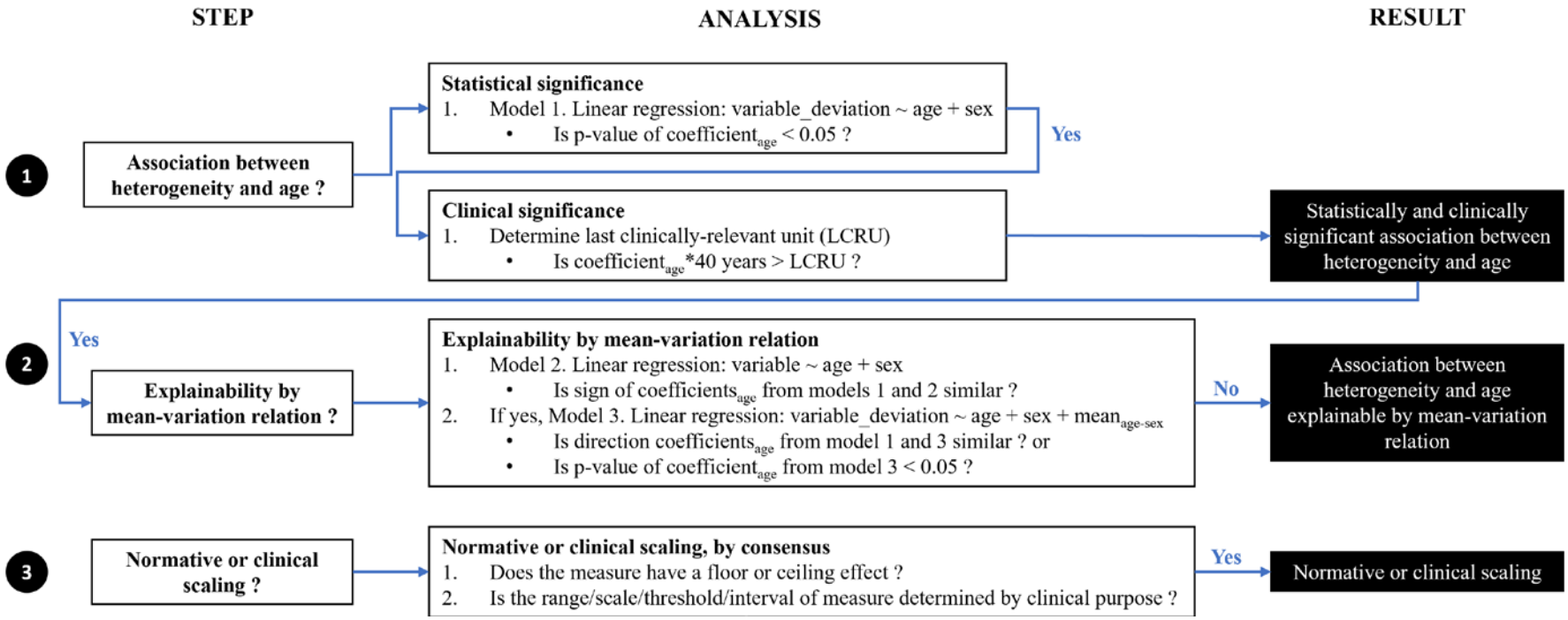
Model 3

Health Characteristic	Combined Significance?	Same Sign Coefficient Models 1 and 2?	Coefficient Dev. Per 10y	p	Coefficient Dev. By Mean (by age, sex)	p	Coefficient Dev. By Sex	p	Different sign or loss of significance after adjustment	Variation Explainable by Mean-Variability Relation
Albumin	No									
Total cholesterol	Yes	No								No
HDL	No									
LDL	Yes	No								No
TSH	Yes	Yes	-0.007	0.414	0.644	0	-0.132	0	Yes	Yes
HbA1C	Yes	Yes	-0.043	0	0.698	0	0.021	0.025	Yes	Yes
Ferritin	Yes	Yes	0.858	0.09	0.374	0	14.619	0	Yes	Yes
Vitamin D	Yes	Yes	-0.376	0.356	0.245	0	0.026	0.965	Yes	Yes
C-reactive protein	No									

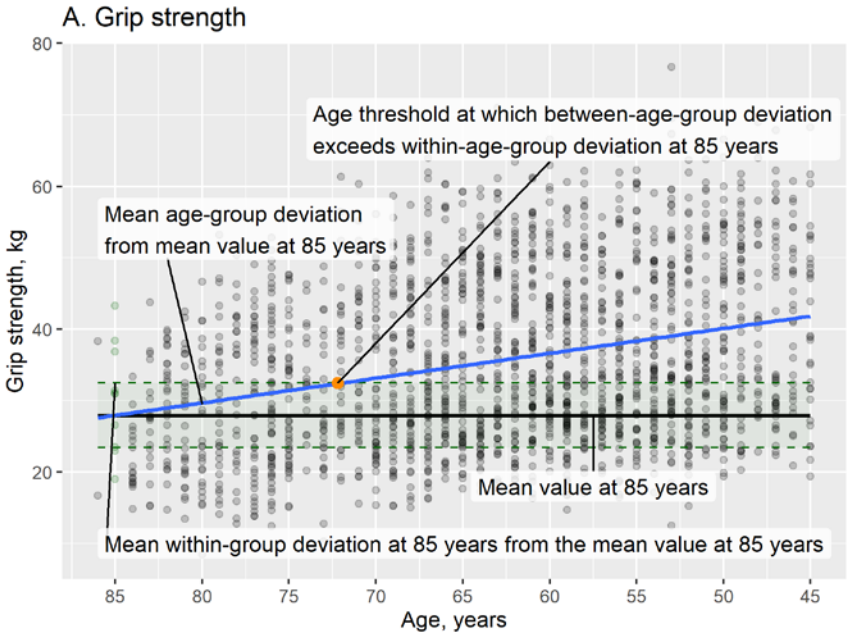
Supplementary Table 4.5. Features and ontology of heterogeneity in health characteristics

Feature	Interpretation
A. Ontology of heterogeneity	Heterogeneity is determined by: - Who is measured ("group/set") - What is measured ("spread") - How it is measured ("measure" and "scaling")
1. Group: who	Group definition determines the boundaries of the group/set that is to be examined
	Group definition determines the set's center point, from which the spread is calculated
2. Spread: what	Heterogeneity is the average distance from the center (as determined by *group*) for an attribute: (i) Baseline health characteristics (ii) Treatment-outcome associations
3. Measure: how	The spread for an attribute in a group is a quantified measure that: (i) Is *scaled*: "natural" *vs* normative or clinical; floor and ceiling effects, skew (ii) Can have a *mean-variation relationship* which can influence heterogeneity
B. Ontology for description and comparison of heterogeneity	Any description or comparative statement about heterogeneity follows from a simultaneous consideration of *group/set + spread + measure*
4. Specificity	Statements about heterogeneity require all 3 features of heterogeneity and cannot be divorced from any
5. Monotonicity	Change in heterogeneity may not be consistent nor be in a consistent direction by the (independent) variable predicting heterogeneity
6. Mean-variation	When comparing heterogeneity between groups, changes in group mean should be considered for potential mean-variation relationships

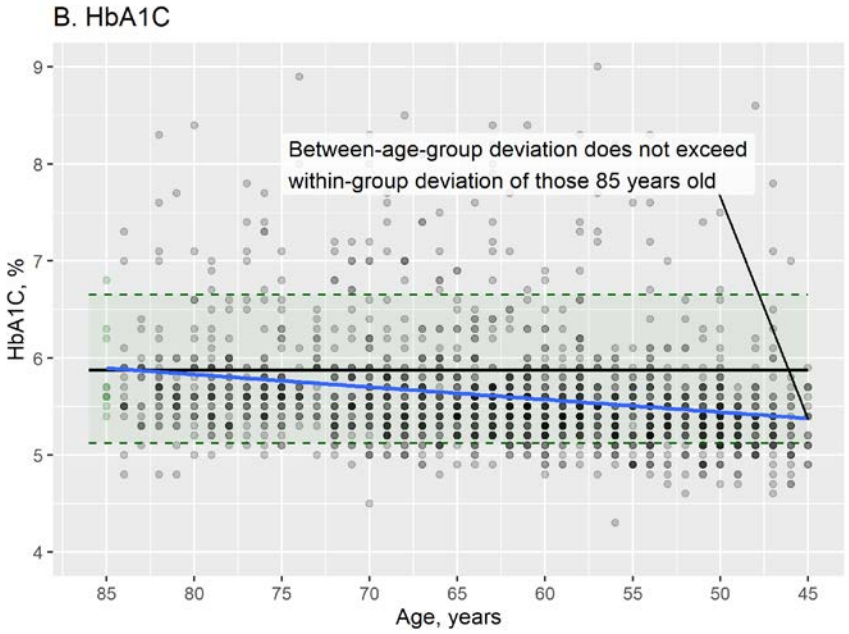
Supplementary Figure 4.1. Flowchart for determining the association between heterogeneity and age, explainability by mean-variation relation and by measurement scaling



Supplementary Figures 4.2. Illustrative scatterplots comparing within-group deviation at 85 years versus between-age-group deviation, for grip strength and HbA1C

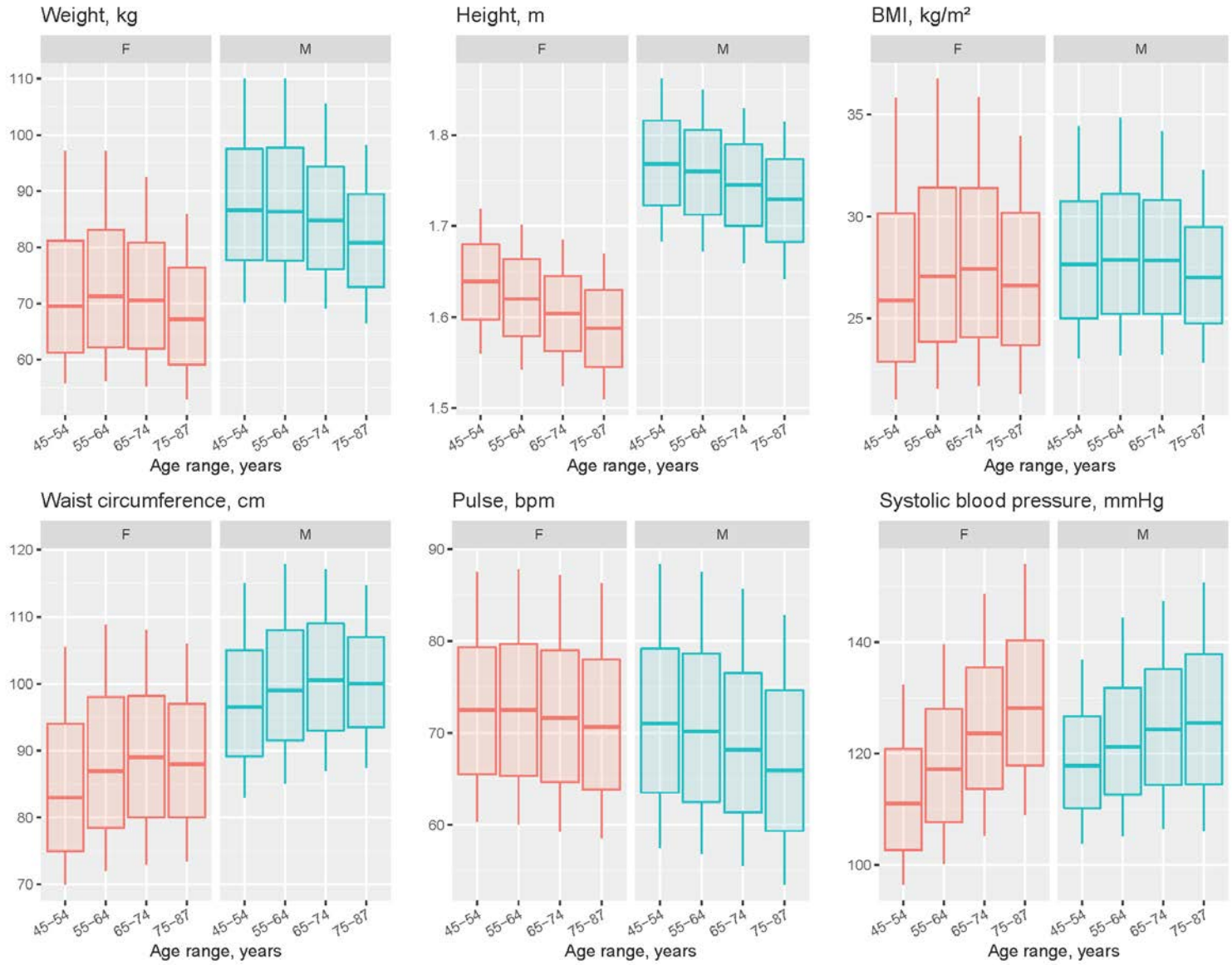


Notes. Note the reversed X-axis.
Above. The green dashed lines are ± the *mean absolute deviation* from the mean value of grip strength within those 85 years old. The slope of the blue line is the age coefficient from a regression of *deviation* of grip strength on age. For grip strength, at approximately 72 years, between-age-group deviation exceeds the deviation within individuals aged 85 years.
Below. In contrast, for HbA1C, between-age-group deviation does not exceed the deviation within individuals aged 85 years. Figure 4.2 in the main text presents the results for the 15 health characteristics where the threshold is crossed between 45 and 85 years.

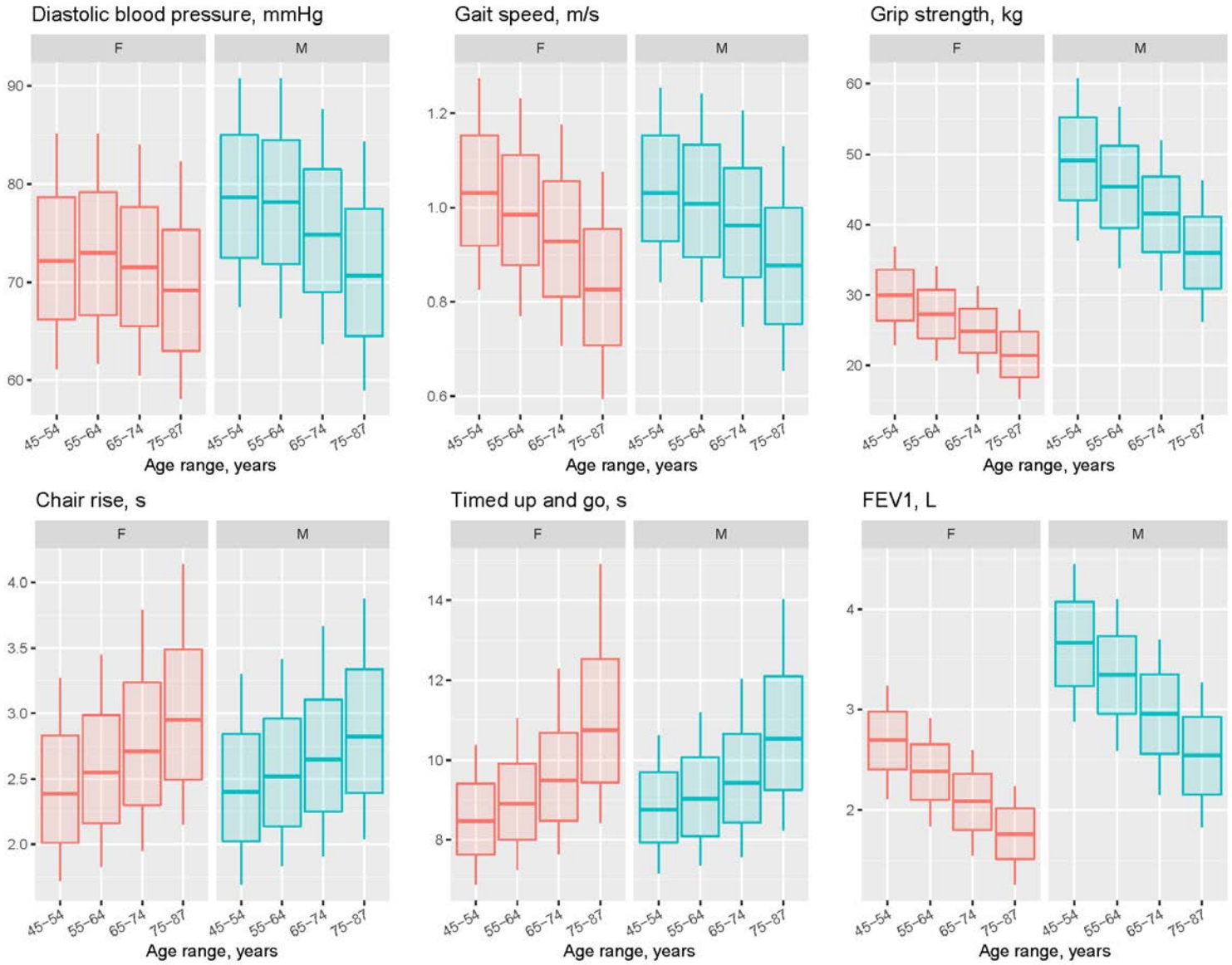


Supplementary Figures 4.3. Boxplots of health characteristics by age and sex

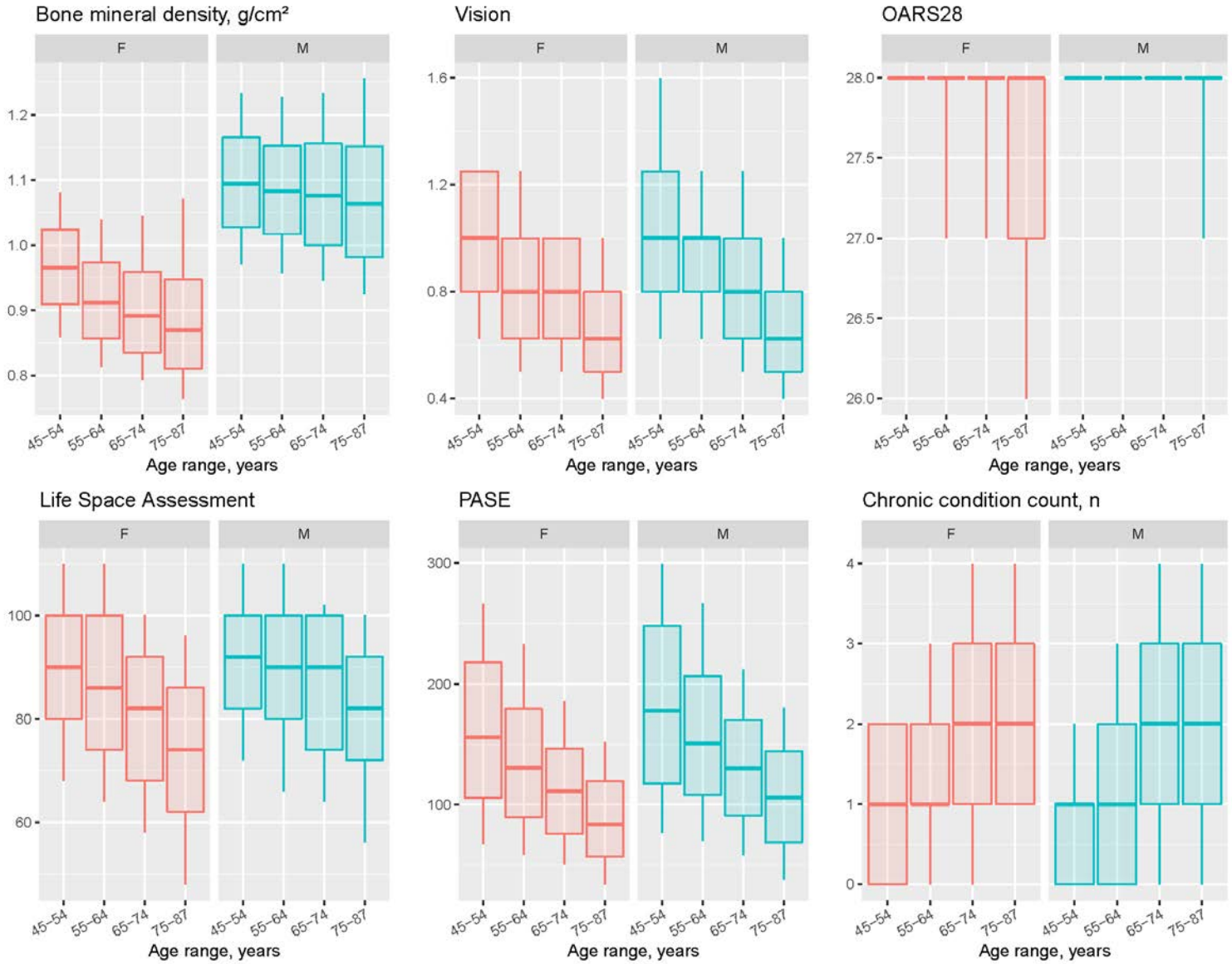
Notes. Middle line in box indicates the median; box indicates the 25th and 75th percentiles; whiskers indicate 10th and 90th percentiles.



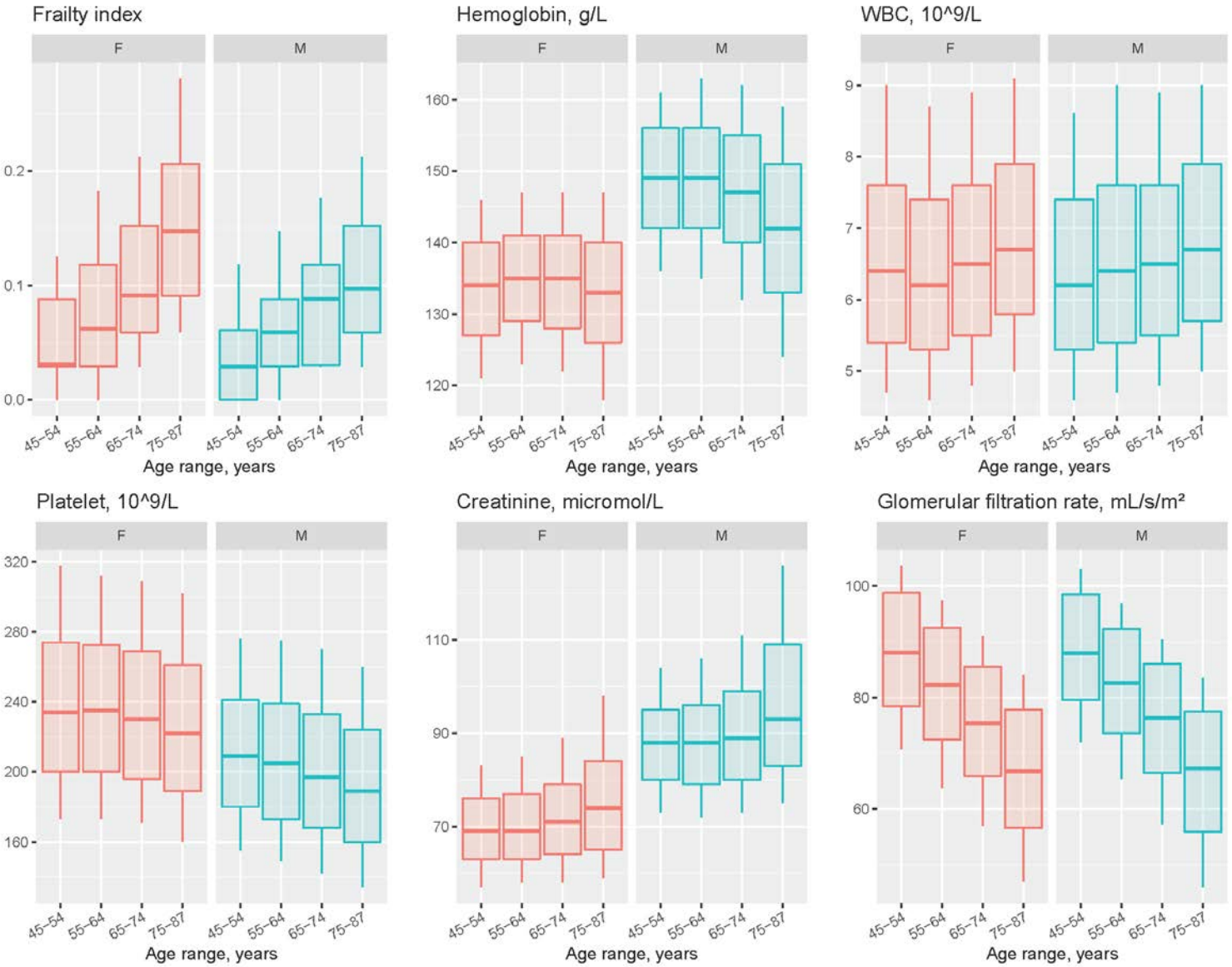
Notes. Middle line in box indicates the median; box indicates the 25th and 75th percentiles; whiskers indicate 10th and 90th percentiles.



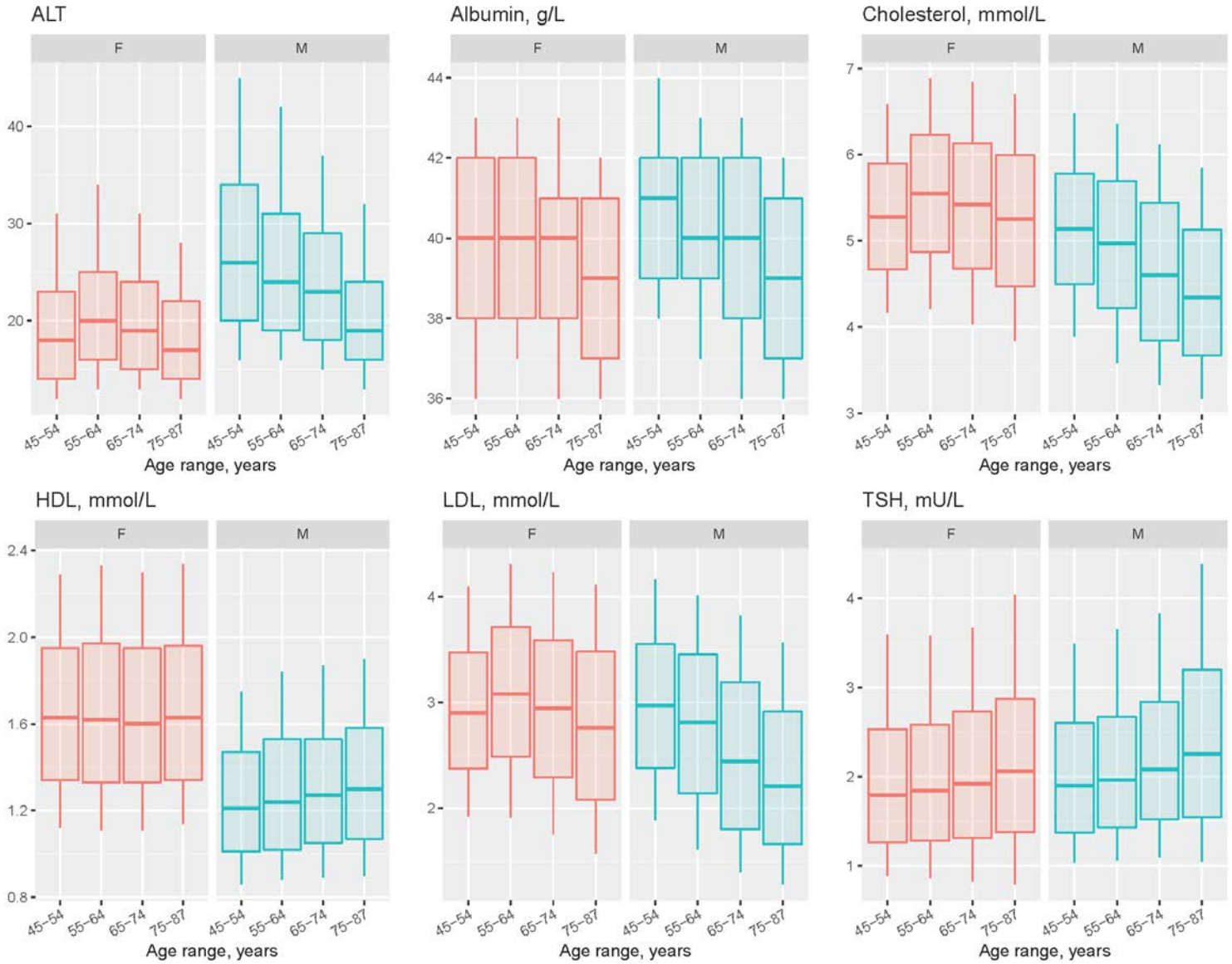
Notes. Middle line in box indicates the median; box indicates the 25th and 75th percentiles; whiskers indicate 10th and 90th percentiles.



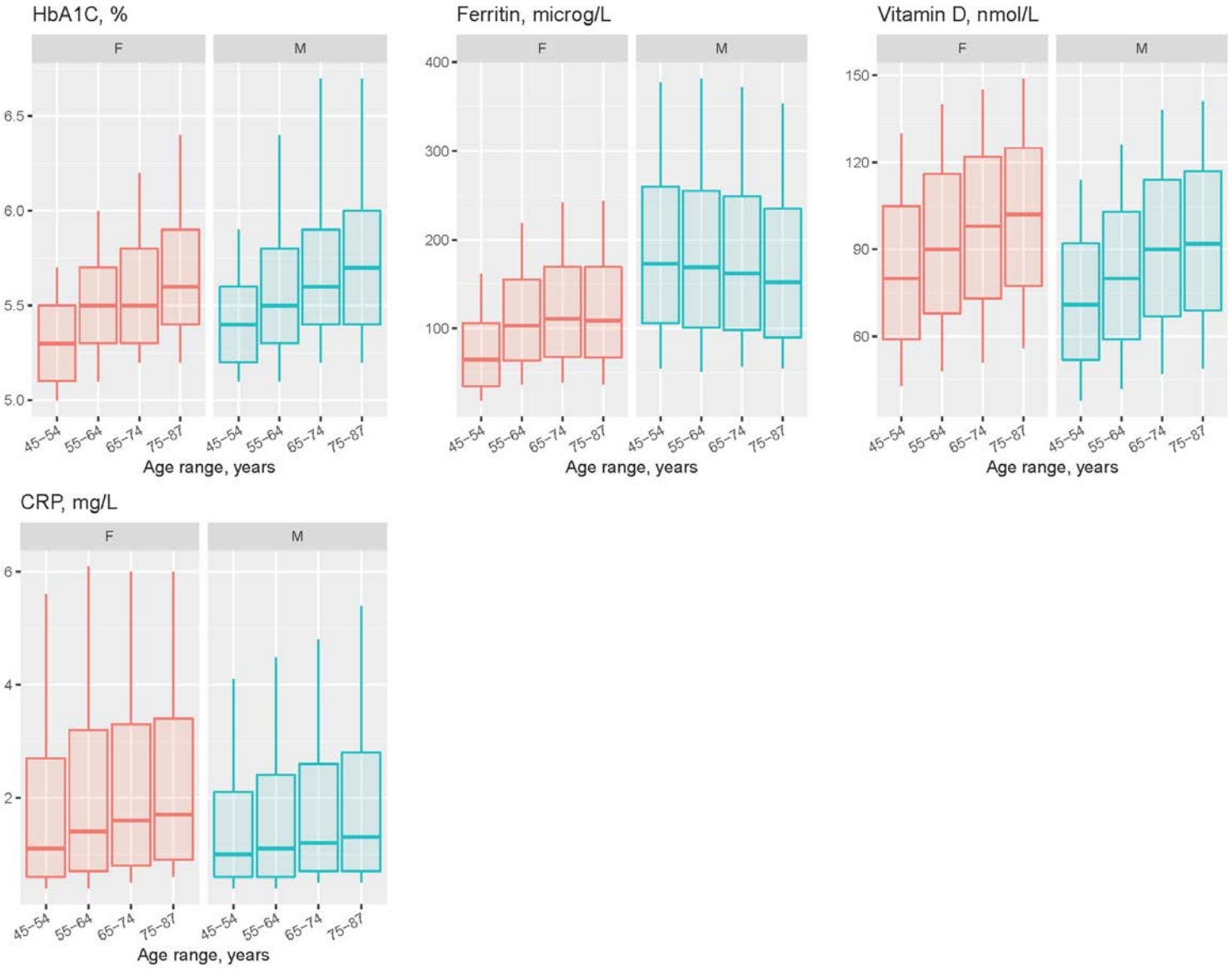
Notes. Middle line in box indicates the median; box indicates the 25th and 75th percentiles; whiskers indicate 10th and 90th percentiles.



Notes. Middle line in box indicates the median; box indicates the 25th and 75th percentiles; whiskers indicate 10th and 90th percentiles.



Notes. Middle line in box indicates the median; box indicates the 25th and 75th percentiles; whiskers indicate 10th and 90th percentiles.



Chapter 5. Manuscript 2: Clinical correlates and implications of the reliability of the frailty index in the Canadian Longitudinal Study on Aging

5.1 PREFACE

Heterogeneity in aging, particularly for age-related health characteristics may provide a suitable substrate for differentiating and thus categorizing older adults. The deficit-accumulation frailty index framework is one influential method to capture age-related heterogeneity due to "frailty" by summing age- and health-related deficits in individuals and dividing by the total number of deficits considered. Frailty indices, as a whole, are remarkably consistent in predicting adverse outcomes in older adults in multiple clinical contexts. However, due to the variability in the deficits composing frailty indices, it is unclear whether existing frailty index implementations are reliable and interchangeable when identifying *individual* older adults as having frailty. Inference about frailty at the individual level may be more stringent than inference about frailty as a construct. For frailty indices to be incorporated into clinical practice, they should systematically identify the same individuals as having frailty, the magnitude of associations between frailty and adverse outcomes should remain stable, and outcome prediction based on a set level of frailty should yield similar risks. In Manuscript 2, I used Monte Carlo methods to simulate 12,000 single studies where frailty indices are computed, described, and examined for associations with mortality. To investigate the reliability—writ large—of frailty indices, the single studies are simulated among the same 12,080 CLSA participants aged 65 years and older and using the same set of 70 health deficits.

***

This manuscript has been published electronically ahead of print in the Journals of Gerontology, Series A: Medical Sciences (https://doi.org/10.1093/gerona/glab161).

5.2 TITLE PAGE

Title

Clinical correlates and implications of the reliability of the frailty index in the Canadian Longitudinal Study on Aging

Short title

Clinical implications of the reliability of the frailty index

Authors and affiliations

Quoc Dinh Nguyen, MD, MA, MPH,[a,b,c] Erica M. Moodie, PhD,[c] Mark R. Keezer, MDCM, PhD,[b,d] Christina Wolfson, PhD[c,e,f]

a. Division of Geriatrics, Centre hospitalier de l'Université de Montréal, Canada;
b. Centre de recherche du Centre hospitalier de l'Université de Montréal, Canada;
c. Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Canada;
d. Departments of Neurosciences & School of Public Health, Université de Montréal, Canada
e. Department of Medicine, McGill University, Canada;
f. Neuroepidemiology Research Unit, Research Institute of the McGill University Health Centre, Canada.

Corresponding author

Quoc Dinh Nguyen
1000, rue Saint-Denis, Montreal,
Quebec, Canada H2X0C1
quoc.nguyen@mail.mcgill.ca

Word counts

Abstract: 250
Main text (from Introduction to Conclusion): 3327
Title page, abstract, and main text (excluding references): 3890
Tables: 3
Figures: 2
References: 38

5.3 ABSTRACT

Background

Deficit-accumulation frailty indices (FIs) are widely used to characterize frailty. FIs vary in number and composition of items; the impact of this variation on reliability and clinical applicability is unknown.

Methods

We simulated 12,000 studies using a set of 70 candidate deficits in 12,080 community-dwelling participants 65 years and older. For each study, we varied the number (5, 10, 15, 25, 35, 45) and composition (random selection) of items defining the FI and calculated descriptive and predictive estimates: frailty score, prevalence, frailty cut-off, mortality odds ratio, predicted probability of mortality for FI=0.28 (prevalence threshold), and FI cut-off predicting 10% mortality over the follow-up. We summarized the estimates' medians and spreads (0.025-0.975 quantiles) by the number of items and calculated intraclass correlation coefficients (ICC).

Results

Medians of frailty scores were 0.11-0.12 with decreasing spreads from 0.04-0.24 to 0.10-0.14 for 5-item and 45-item FIs. The median cut-offs identifying 15% as frail was 0.19-0.20 and stable; the spreads decreased with more items. However, medians and spreads for the prevalence of frailty (medians: 11% to 3%), mortality odds ratio (medians:1.24 to 2.19), predicted probability of mortality (medians: 8% to 17%), and FI cut-off predicting 10% mortality (medians: 0.38 to 0.20) varied markedly. ICC increased from 0.19 (5-item FIs) to 0.84 (45-item FIs).

Conclusions

Variability in the number and composition of items of individual FIs strongly influences their reliability. Estimates using FIs may not be sufficiently stable for generalizing results or direct application. We propose avenues to improve the development, reporting, and interpretation of FIs.

Keywords

CLSA, measurement error, psychometrics, reliability, regression dilution

5.4 INTRODUCTION

Frailty, conceptualized as a state of increased vulnerability to stressors, is often used to characterize heterogeneity in aging[1] in research and clinical practice.[2,3] While there is currently no consensus definition for frailty, the deficit accumulation frailty index (FI)[4,5] is one of two major frameworks underpinning the current understanding of frailty, along with the Fried frailty phenotype.[6–8] The deficit accumulation frailty index conceptualizes frailty as the accumulation of age-related deficits, which is quantified using a standard procedure that calculates the proportion of deficits across a selection of 35 or more items measuring health states in multiple age-related domains (e.g., physical function, medical conditions).[9] Since first described in 2001,[4] numerous other deficit accumulation frailty indices have been developed using different sources of data (e.g., comprehensive geriatric assessments,[10] electronic medical records,[11] laboratory values,[12] trial,[13] or registry[14] data), number of items (from 5 to 92),[15,16] from different domains,[17] and using various cut-offs.[18,19]

To incorporate frailty measures into clinical practice, frailty scores or frailty status measures should identify the same individuals as frail from one context or study to another. In a recent review of the reliability of 35 frailty instruments, frailty indices have shown the highest agreement among frailty frameworks.[19] However, only six well-constructed FIs were included in the analysis, whereas the myriad implementations of FIs currently in use may not demonstrate the same reliability. Although coding items as dichotomous or ordinal does not affect the performance of FIs,[20] the influence of variation in the number and the composition of items on reliability is unknown. Moreover, classical measurement statistics of reliability, such as intraclass correlation coefficients, kappa, or standard error of measurement,[21,22] may not translate easily to clinical correlates and implications when incorporating frailty in practice. For instance, how reliability

influences estimates such as prevalence,[19,23] association with outcomes, cut-offs categorizing frailty status (frail vs. non-frail) has not been well described. Clinical decision-making at the individual level may require more stringent conditions than population-level inferences.[24,25]

In this study, we investigated the reliability of frailty indices and the stability of estimates when computed in a single community-dwelling older adult population as represented by participants from the Canadian Longitudinal Study on Aging (CLSA). We simulated 12,000 studies in which we varied the number of items and the specific composition of individual frailty indices to describe the "in vivo" implications of frailty indices for developing, reporting, and interpreting studies using the frailty index.

5.5 METHODS

5.5.1 Cohort

We used baseline cross-sectional data from the CLSA which enrolled a nationally representative sample of over 51,000 participants aged 45 to 85 years, at baseline, into a telephone-only cohort or an in-person cohort (Comprehensive) assessed from 2012 to 2015.[26,27] As frailty is an age-related concept and is typically measured in older adult populations, we restricted our cohort to adults 65 years and older with mortality data as of July 2019, and to the Comprehensive cohort since data from physical assessments were required. Exclusion criteria for the CLSA were people living in the Canadian territories, on a First Nations reserve, or in institutions, being full-time members of the Armed Forces, having cognitive impairment (as assessed by interviewers), and being unable to respond in French or English. There were 30,097 CLSA participants in the Comprehensive cohort of whom 12,080 met the inclusion criteria for our analyses.

5.5.2 Methodological framework

The methodological framework for our study is depicted in Figure 5.1. To investigate the inter-index reliability of the frailty index across multiple configurations, we generated and compared a total of 12,000 iterations, each representing a potential individual study. Each iteration (or individual study) applied a unique definition of the frailty index by varying (i) the *number* of health deficits composing the frailty index (we chose six configurations: 5, 10, 15, 25, 35, and 45 items, with 2000 iterations per configuration for a total of 12,000 single studies); and (ii) by randomly selecting the *specific* deficits composing each. Configurations with few items (i.e., 5, 10, and 15) were included since two highly cited and influential deficit-accumulation derived frailty indices in usage include 5 and 11 items.[14,15,28]

5.5.3 Measure, outcome, and analyses at the single-study level

At the individual-study level, we created frailty indices using the standard procedure outlined by Searle et al.[9] Briefly, the frailty index is calculated as the proportion of health deficits in an individual, with health deficits satisfying the following conditions for a single time-point study: (i) deficits must be associated with health status, (ii) their prevalence must generally increase with age, (iii) deficits should not saturate too early, and (iv) they should cover a range of systems as a group. We considered a collection of 70 health deficits, across 10 domains, as the basis for our frailty index definitions. Deficits were chosen to map closely to those in the most cited frailty index proposed by Rockwood et al. in 2005 using items from the Canadian Study of Health and Aging (CSHA).[5] Supplementary Table 5.1 reports the health deficits and domains, cut-offs for categorizing continuous deficits, missing data for deficits, and their mapping to CSHA items. Individual deficits were assessed in-person and by self-report; for details, see the CLSA Protocol.[27] The CLSA updated the mortality data in July 2019 (median follow-up: 5.6 years, IQR: 1.4 years).

The mortality data came from the next of kin contacting CLSA directly, identification of death at the time of follow-up, and from linkage to provincial vital statistics. Because date of death was not available, mortality status was determined for all participants in July 2019.

Analyses

Within each single study, we first computed descriptive estimates: (1) mean population frailty score (in the 12,080 participants), (2) prevalence of frailty (as defined as a FI above 0.28, the mean of cut-offs for the frailty indices in our scoping review of 150 articles on frailty, 35 of which used frailty indices [range of cut-offs = 0.18-0.41; forthcoming]), (3) the FI threshold that categorizes 15% of the population as frail (i.e., defined as the 85th percentile of FI).[29] We then computed predictive estimates between the frailty index and mortality using logistic regression: (4) the odds ratio (OR) for a 0.1 increase in the FI, (5) the average predicted risk of mortality over the follow-up period for a FI of 0.28, and (6) the FI cut-off predicting a 10% risk of mortality. Together, these 6 estimates account for common descriptive and clinical usages of frailty: describing a population, identifying those with frailty, estimating the association between frailty and an outcome, predicting risk for an outcome, and identifying a risk threshold for decision-making.

5.5.4 Analyses and outcome of interest at the comparative level

We compared the 6 different configurations of the frailty index using 2000 iterations (i.e., "individual studies"). At the comparative level, our primary outcome of interest was the stability of each descriptive and predictive estimate, as measured by the median and the 2.5 and 97.5 percentiles. We also computed classical reliability statistics for each configuration: the intraclass correlation coefficient for agreement ($ICC_A = \sigma^2_{individuals} / [\sigma^2_{individuals} + \sigma^2_{frailty\ indices} + \sigma^2_{residual}]$) and the standard error of measurement for agreement, $SEm_A = \sqrt{(\sigma^2_{frailty\ indices} + \sigma^2_{residual})}$, which scales the ICC_A on the scale of the frailty index.[30] Using ICC in this situation is analogous to

assessing inter-rater reliability where "raters" are the randomly defined FIs. Following guidelines presented by Nunnally, a threshold of 0.9 was considered acceptable for ICC_A in this clinical context.[22,25] As an exploratory analysis on the impact of domain coverage of frailty indices, we repeated the stability of estimates analyses above, stratifying by the number of domains included in 10, 25, and 45-item frailty index configurations (as determined by the inclusion of at least one item from that domain). We calculated the ICC_A for all configurations and number of domains. Analyses were performed using R 4.0.3 (R Foundation).

5.6 RESULTS

Among the 12,080 community-dwelling participants, the mean (SD) age was 73.0 (5.7) years, and 6097 (50.5%) were male. Impairment in activities of daily living (ADL, n = 495 [4.1%]) and instrumental ADL (iADL, n = 982 [8.2%]) was relatively infrequent. Self-reported hypertension (n = 5937 [49.3%]) and osteoarthritis (n = 4167 [34.5%]) were the most prevalent conditions. Overall, participants had good physical performance measures: mean time for single chair rise was 2.9 (0.9) seconds and mean gait speed was 0.9 (0.2) m/s. Table 5.1 reports the baseline characteristics of our study sample. Between the baseline and mortality assessment (median follow-up of 5.6 years, IQR = 1.4]), 762 (6.3%) participants were known to have died.

5.6.1 Reliability and stability of frailty indices measurements and estimates

Complete results for the reliability of frailty indices and stability of estimates are reported in Table 5.2. Figure 5.2 provides a graphical summary of results for the stability and distribution of estimates.

Descriptive uses of the frailty index

The medians of the mean frailty index for all 6 configurations (5, 10, 15, 25, 35, and 45 items) was 0.11-0.12 with decreasing spreads between the 2.5 and 97.5 percentiles from 0.04-0.24 to 0.10-

0.14 for 5-item FIs and 45-item FIs, respectively. Figure 5.2A shows the distribution of FIs by configurations and the increasing narrowness and smoothness of distributions as the number of items increases. The prevalence of frailty varied widely between configurations: the median prevalence was 0.11 for 5-item FIs but decreased to 0.03 for 45-item FIs; within each configuration, the 0.025 to 0.975 quantiles spreads of prevalence ranged from 0.02-0.33 to 0.02-0.06. The medians of the frailty index cut-off identifying 15% of participants as frail was stable at 0.19-0.20; the spreads of this cut-off progressively decreased from 0.20-0.40 to 0.16-0.22.

Predictive uses of the frailty index and measurement statistics

The medians of the odds ratio from regressions of mortality on the frailty index varied, increasing from 1.24 (95% CI: 1.02, 1.43) for 5-item FIs to 2.19 (1.97, 2.47) for 45-item FIs. Likewise, the predicted probabilities of mortality over the follow-up period in individuals with a frailty index of 0.28 increased from 0.08 (0.06, 0.12) up to 0.17 (0.14, 0.21). Both the medians of the frailty index and spreads identifying a 10% risk of mortality decreased from 0.38 (0.18, 1.23) for 5-item FIs to 0.20 (0.18, 0.23) for 35-item FIs. The intraclass correlation coefficients were higher as the number of items included increased, starting at 0.19 (0.18, 0.19) and reaching 0.84 (0.84, 0.85) for 45-item FIs. The standard error of measurement for agreement was substantial at 0.13 (0.13, 0.13) for 5-item FIs and decreased progressively to 0.03 (0.03, 0.03) for 45-item FIs.

Reliability and stability by number of domains

Exploratory analysis results for frailty indices stratified by the number of domains are presented in Supplementary Figure 5.1 for the stability of estimates and in Supplementary Figure 5.2 for ICC_A. Although there were differences in the stability of estimates for 10-item FIs by the number of domains included, the variability between the number of items included outweighed the

variability of the number of domains covered. The ICC_A followed a similar trend, where most of the differences in reliability was due to the total number of items.

5.7 Discussion

In this study of a single older adult population, we generated frailty index definitions based on the same set of 70 health deficits but varied both the number of deficits and the specific deficits considered. We show notable variation in the descriptive and predictive estimates computed from frailty indices, e.g., mean score, prevalence, categorical frailty status, odds ratio, frailty cut-off and risk prediction. The instability of estimates was twofold: *within* configurations and *between* configurations. Within each configuration using the same number of deficits, the instability of estimates is observed as the spread around each median value. When comparing between configurations, we demonstrate additional variability in the median values themselves for the prevalence (from 5-item FIs to 45-item FIs: 0.11 to 0.03), the odds ratio (1.24 to 2.19), the frailty indices predicting 10% mortality (0.08 to 0.17), and the estimated risk of mortality for a FI of 0.28 (0.38 to 0.20).

Previous work has investigated the psychometric properties and comparability of major existing frailty frameworks,[19,31–34] as well as of frailty indices specifically.[16,20] Although variation in prevalence and low interchangeability between frameworks is attributable to different underlying concepts of frailty [19], frailty indices report the highest agreement among frailty frameworks. As a group, frailty indices have good criterion and construct validity,[16] but misclassification and decreased predictive accuracy may occur with the reduction of domains composing frailty indices.[35] Our findings add to this body of research by focusing on (i) the reliability of the frailty indices and stability of estimates, (ii) the clinical implications of the

reliability of frailty indices, specifically for patient care, and (iii) the differences between their population and individual interpretations of studies.

5.7.1 Reliability of frailty indices, stability of estimates, and regression dilution

Psychometrically, an ideal frailty measurement, used repeatedly in a short timeframe where health status is stable, should yield similar scores and identify, as frail, the same individuals. Intraclass correlation for agreement, the ratio of the variation *between individuals* and the *overall* variation (including variation due to measurement), is a classical measure of reliability. We found that when using a 10-item frailty index ($ICC_A = 0.34$), only 34% of the variation in frailty scores was due to differences between individuals. As expected, we show that reliability improves as the number of items included in computing frailty indices increased. Yet, even with 35 and 45 items, the ICC_A is 0.75 and 0.84, respectively, which is less than the 0.90 threshold for clinical decision-making recommended by Nunally.[22,25]

The reliability of frailty indices directly informs prevalence estimates: the lower the reliability, the wider the spread of prevalence estimates. Interestingly, the median prevalence also varied by the number of items included (from 0.11 to 0.03) due to the non-smooth distribution of frailty scores in relation to the location of the frailty cut-off, as shown in Figure 5.2A. Varying frailty prevalence has been attributed to differing frailty operationalizations, cut-offs, and populations under study; we show that variation in prevalence may also be due to measurement issues, even within the frailty index framework using a single cut-off.

Using a small number of items increases measurement error which has important consequences on predictive estimates. This is illustrated in Table 5.2 showing that the magnitude of the association between frailty and mortality lessens as the number of items decreases. When

exposures are mismeasured, as is the case in FIs using a lower number of items, the well-known phenomenon of regression dilution can occur, whereby mismeasurement biases the true association toward the null.[36,37] In a seminal paper outlining the standard procedure to create a frailty index, Searle et al. specify that at least 30-40 items should be included and that estimates are unstable when the number of items (~10) is small.[9] This cautionary recommendation has not been completely heeded; according to our scoping review (forthcoming), the median number of items included was 32 in a random subset of 35 studies using frailty indices published in 2017 and 2018, of which 13 used the 5 or 11-item modified frailty index.[14,15,28] Whereas previous work has highlighted the role of domain coverage in optimizing the predictive accuracy of frailty indices,[35] our findings suggest that the total number of items included, rather than domain coverage, is more influential for the reliability of frailty indices.

5.7.2 Clinical implications of the reliability of frailty indices

The reliability of frailty indices has direct implications on their application to clinical practice. Our findings indicate first that frailty indices, especially those with a small number of items, can vary in scores attributed to each individual from one constructed index to another. Identifying frailty status as a consistent basis for decision-making is challenging due to the large standard error of measurement, from 0.13 for 5 items to 0.03 for 45 items, a substantial variation when considering the distribution of frailty scores. Second, in addition to reliability issues, the accuracy of the frailty index, continuous or categorical, to predict outcomes such as mortality, may be decreased due to regression dilution. Third, establishing cut-offs of the frailty index may yield unexpected results because of the non-smooth distribution of frailty when measured using fewer than 25 items. Fourth, generalizing or transporting results between frailty indices or to clinical settings can only be done with caution since the average “difficulty” of items comprising each frailty index may

vary markedly. In a single population, the mean frailty score may be 0.10 for one frailty index and 0.20 in another, without the possibility of identifying where each frailty index is anchored relative to the other in the underlying (latent) frailty continuum. Comparatively, other frailty instruments, such as the Clinical Frailty Scale[5] and the components of the Fried physical frailty phenotype,[7] have absolute anchors by way of descriptive rubrics or explicit cut-offs, respectively. It may be useful to note that the clinical implications we describe are not unique to the frailty index: other constructs in geriatric practice and aging research (e.g., multimorbidity) do not have a well-defined set of components, cut-offs, or anchoring.

5.7.3 Differences between population and individual interpretation of frailty indices

Although we focused on the clinical application of frailty indices, their usage in research to draw inferences at the population level may not face the same limitations,[24] Psychometric issues are critical when translating research findings to individuals in clinical practice.[22] However, the lack of reliable anchoring of frailty indices does not invalidate their consistent associations with adverse outcomes. At the *population level*, frailty is indeed associated with falls, progression of disability, and mortality;[3] but what specific level of frailty can reliability predict falls and the initiation of fall prevention program at the *individual level* remains undetermined and varies between different constructed indices. A continuum in applicability exists whereby the requirements for clinical applicability is higher than that for research and population-level interpretation. In light of our findings, Table 5.3 presents recommendations for clinical usage and research of frailty indices, when devising, reporting, and interpreting frailty indices.

5.7.4 Limitations

Our study has a few important limitations that should be recognized. First, measures of reliability are dependent on the sample in which they are assessed. We included older adults 65 years and

older from the CLSA which excluded institutionalized and cognitively impaired adults at baseline; the findings from these analyses may not be transportable to a younger or a less healthy population. Second, our results may be sensitive to the 70 candidate items and cut-offs we selected as a basis for our frailty indices; however, we chose items to map to the original CSHA items,[5] using standard or first-quintile cut-offs as recommended.[9] Third, we examined frailty indices with a number of items below the recommended 30-35 which lowers reliability; nonetheless, a large proportion of frailty indices are defined using a small number of items.[12,14,15] Fourth, because we used the same 70 candidate items and cut-offs for our frailty indices, our analysis may actually *overestimate* the reliability of frailty indices currently in usage which may use more diverse items and cut-offs. As electronic medical records may allow the calculation of more reliable frailty indices due to greater availability of items, their cut-off and "level of difficulty" will be increasingly important to consider. Finally, because the date of death was not available, associations and predictions between frailty and mortality using logistic regressions should be interpreted with caution, although comparisons remain valid between the predictive estimates of each configuration.

5.8 CONCLUSION

By simulating 12,000 individual studies, we show that variability in the number and the composition of items of individual frailty indices strongly influences their overall reliability. Descriptive and predictive estimates using frailty indices may not be sufficiently stable for generalizing results or for direct application to clinical practice. Although reliability improves as the number of items is increased, we propose further avenues to improve the development, reporting, and interpretation of frailty indices.

5.9 Acknowledgements

This research was made possible using the data/biospecimens collected by the Canadian Longitudinal Study on Aging (CLSA). Funding for the Canadian Longitudinal Study on Aging (CLSA) is provided by the Government of Canada through the Canadian Institutes of Health Research (CIHR) under grant reference: LSA 94473 and the Canada Foundation for Innovation. This research has been conducted using the Baseline CLSA Comprehensive V4.0 dataset and the Follow-up 1 Comprehensive V1.0 dataset, under Application Number 190209. The CLSA is led by Drs. Parminder Raina, Christina Wolfson and Susan Kirkland. The opinions expressed in this manuscript are the author's own and do not reflect the views of the Canadian Longitudinal Study on Aging.

Author contributions

QDN, EMM, CW designed the study. QDN analyzed data and drafted the manuscript. All authors contributed significantly to the content, critically reviewed, and approved the final manuscript for publication.

Conflict of interests

None.

5.10 REFERENCES

1. Nguyen QD, Moodie EM, Forget M-F, Desmarais P, Keezer MR, Wolfson C. Health heterogeneity in older adults: exploration in the Canadian Longitudinal Study on Aging. J Am Geriatr Soc. 2021;69(3):678-687. doi:https://doi.org/10.1111/jgs.16919

2. Dent E, Kowal P, Hoogendijk EO. Frailty measurement in research and clinical practice: a review. *Eur J Intern Med*. 2016;31:3-10. doi:10.1016/j.ejim.2016.03.007

3. Hoogendijk EO, Afilalo J, Ensrud KE, Kowal P, Onder G, Fried LP. Series Frailty 1 Frailty : implications for clinical practice and public health. *Lancet*. 2019;394(10206):1365-1375. doi:10.1016/S0140-6736(19)31786-6

4. Mitnitski AB, Mogilner AJ, Rockwood K. Accumulation of deficits as a proxy measure of aging. *Sci World*. 2001;1:323-336. doi:10.1100/tsw.2001.58

5. Rockwood K, Song X, MacKnight C. A global clinical measure of fitness and frailty. *Can Med Assoc J*. 2005;173(5):489-495.

6. Buta BJ, Walston JD, Godino JG, et al. Frailty assessment instruments: systematic characterization of the uses and contexts of highly-cited instruments. *Ageing Res Rev*. 2016;26:53-61. doi:10.1016/j.arr.2015.12.003

7. Fried LP, Tangen CM, Walston J, et al. Frailty in older adults: evidence for a phenotype. *J Gerontol A Biol Sci Med Sci*. 2001;56(3):808-813. doi:10.1093/gerona/56.3.M146

8. de Vries NM, Staal JB, van Ravensberg CD, Hobbelen JSM, Rikkert MGMO, der Sanden MWGN. Outcome instruments to measure frailty: a systematic review. *Ageing Res Rev*. 2011;10(1):104-114. doi:10.1016/j.arr.2010.09.001

9. Searle SD, Mitnitski A, Gahbauer EA, Gill TM, Rockwood K. A standard procedure for creating a frailty index. *BMC Geriatr*. 2008;8(1):24. doi:10.1186/1471-2318-8-24

10. Jones DM, Song X, Rockwood K. Operationalizing a Frailty Index from a Standardized Comprehensive Geriatric Assessment. *J Am Geriatr Soc*. 2004;52(11):1929-1933. doi:10.1111/j.1532-5415.2004.52521.x

11. Clegg A, Bates C, Young J, et al. Development and validation of an electronic frailty index

using routine primary care electronic health record data. *Age Ageing*. 2016;45(3):353-360. doi:10.1093/ageing/afw039

12. Howlett SE, Rockwood MRH, Mitnitski A, Rockwood K. Standard laboratory tests to identify older adults at increased risk of death. *BMC Med*. 2014;12:171. doi:10.1186/s12916-014-0171-9

13. Pajewski NM, Williamson JD, Applegate WB, et al. Characterizing Frailty Status in the Systolic Blood Pressure Intervention Trial. *J Gerontol A Biol Sci Med Sci*. 2016;71(5):649-655. doi:10.1093/gerona/glv228

14. Farhat JS, Velanovich V, Falvo AJ, et al. Are the frail destined to fail? Frailty index as predictor of surgical morbidity and mortality in the elderly. *J Trauma Acute Care Surg*. 2012;72(6):1526-1530; discussion 1530-1. doi:10.1097/TA.0b013e3182542fab

15. Subramaniam S, Aalberg JJ, Soriano RP, Divino CM. New 5-Factor Modified Frailty Index Using American College of Surgeons NSQIP Data. *J Am Coll Surg*. 2018;226(2):173-181.e8. doi:10.1016/j.jamcollsurg.2017.11.005

16. Drubbel I, Numans ME, Kranenburg G, Bleijenberg N, De Wit NJ, Schuurmans MJ. Screening for frailty in primary care: a systematic review of the psychometric properties of the frailty index in community-dwelling older people. *BMC Geriatr*. 2014;14(1):1-13. doi:10.1186/1471-2318-14-27

17. Sternberg SA, Wershof Schwartz A, Karunananthan S, Bergman H, Mark Clarfield A. The identification of frailty: a systematic literature review. *J Am Geriatr Soc*. 2011;59(11):2129-2138. doi:10.1111/j.1532-5415.2011.03597.x

18. Romero-Ortuno R. An alternative method for Frailty Index cut-off points to define frailty categories. *Eur Geriatr Med*. 2013;4(5):299-303. doi:10.1016/j.eurger.2013.06.005

19. Aguayo GA, Donneau AF, Vaillant MT, et al. Agreement between 35 published frailty scores in the general population. *Am J Epidemiol*. 2017;186(4):420-434. doi:10.1093/aje/kwx061

20. Peña FG, Theou O, Wallace L, et al. Comparison of alternate scoring of variables on the performance of the frailty index. *BMC Geriatr*. 2014;14(1). doi:10.1186/1471-2318-14-25

21. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international Delphi study. *Qual Life Res*. 2010;19(4):539-549. doi:10.1007/s11136-010-9606-8

22. Streiner DL, Norman GR, Cairney J. *Health Measurement Scales A Practical Guide to Their Development and Use*. 5th ed. Oxford, UK: Oxford University Press; 2014. doi:10.1093/med/9780199685219.001.0001

23. Collard RM, Boter H, Schoevers RA, Oude Voshaar RC. Prevalence of frailty in community-dwelling older persons: a systematic review. *J Am Geriatr Soc*. 2012;60(8):1487-1492. doi:10.1111/j.1532-5415.2012.04054.x

24. Henderson R, Keiding N. Individual survival time prediction using statistical models. *J Med Ethics*. 2005;31(12):703-706. doi:10.1136/jme.2005.012427

25. Nunnally JC. *Psychometric Theory*. 2nd ed. McGraw-Hill; 1978.

26. Raina P, Wolfson C, Kirkland S, et al. Cohort Profile: The Canadian Longitudinal Study on Aging (CLSA). *Int J Epidemiol*. 2019;48(6):1752-1753j. doi:10.1093/ije/dyz173

27. Raina PS, Wolfson C, Kirkland SA. Canadian Longitudinal Study on Aging - Protocol. https://clsa-elcv.ca/doc/511. Accessed November 26, 2019.

28. Velanovich V, Antoine H, Swartz A, Peters D, Rubinfeld I. Accumulating deficits model of frailty and postoperative mortality and morbidity: its application to a national database. *J Surg Res*. 2013;183(1):104-110. doi:10.1016/j.jss.2013.01.021

29. Bandeen-Roche K, Seplaki CL, Huang J, et al. Frailty in older adults: A nationally representative profile in the United States. *J Gerontol A Biol Sci Med Sci*. 2015;70(11):1427-1434. doi:10.1093/gerona/glv133

30. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. *Measurement in Medicine: A Practical Guide*. Cambridge: Cambridge University Press; 2011. doi:DOI: 10.1017/CBO9780511996214

31. Walston JD, Bandeen-Roche K. Frailty: a tale of two concepts. *BMC Med*. 2015;13:185.

doi:10.1186/s12916-015-0420-6

32. Rockwood K, Andrew M, Mitnitski A. A comparison of two approaches to measuring frailty in elderly people. *J Gerontol A Biol Sci Med Sci*. 2007;62(7):738-743. doi:10.1093/gerona/62.7.738

33. Rockwood K, Blodgett JM, Theou O, et al. A Frailty Index Based on Deficit Accumulation Quantifies Mortality Risk in Humans and in Mice. *Sci Rep*. 2017;7:1-10. doi:10.1038/srep43068

34. Theou O, Brothers TD, Mitnitski A, Rockwood K. Operationalization of Frailty Using Eight Commonly Used Scales and Comparison of Their Ability to Predict All-Cause Mortality. *J Am Geriatr Soc*. 2013;61(9):1537-1551. doi:10.1111/jgs.12420

35. Shi SM, McCarthy EP, Mitchell S, Kim DH. Changes in Predictive Performance of a Frailty Index with Availability of Clinical Domains. *J Am Geriatr Soc*. 2020;68(8):1771-1777. doi:10.1111/jgs.16436

36. Keogh RH, White IR. A toolkit for measurement error correction, with a focus on nutritional epidemiology. *Stat Med*. 2014;33(12):2137-2155. doi:10.1002/sim.6095

37. Hutcheon JA, Chiolero A, Hanley JA. Random measurement error and regression dilution bias. *BMJ*. 2010;340:c2289.

38. Lord FM, Novick MR, Birnbaum A. *Statistical Theories of Mental Test Scores.* Oxford, England: Addison-Wesley; 1968.

5.11 TABLES

Table 5.1. Baseline characteristics of participants 65 years and older of the Canadian Longitudinal Study on Aging (n = 12,080)

Age, mean (SD)	73.0 (5.7)
Male sex (%)	6097 (50.5)
White race/ethnicity (%)	11673 (96.6)
Married or living with partner (%)	7051 (61.7)
Living location (%)	
House	8760 (72.6)
Apartment or condominium	3126 (25.9)
Seniors' housing	132 (1.1)
Other	53 (0.4)
ADL impairment (%)	495 (4.1)
IADL impairment (%)	982 (8.2)
Chronic conditions (%)	
Hypertension	5937 (49.3)
Diabetes	2624 (21.8)
Heart disease	2772 (23.0)
Stroke or transient ischemic attack	918 (7.6)
Lung disease	2017 (16.7)
Kidney disease	499 (4.1)
Thyroid disease	2156 (18.2)
Osteoarthritis	4167 (34.5)
Osteoporosis	1701 (14.3)
Cancer	2803 (23.3)
Anxiety or depression	2053 (17.0)
Physical performance measures, mean (SD)	
Grip strength (kg)	31.7 (10.6)
Chair rise time (s)	2.9 (0.9)
Gait speed (m/s)	0.9 (0.2)

Table 5.2. Reliability and stability of frailty indices measurements and estimates for descriptive and predictive uses, by the number of items in 6 configurations

	Number of items in frailty index in configuration (2000 iterations for each configuration)					
	5	**10**	**15**	**25**	**35**	**45**
Descriptive uses of the frailty index, median (0.025 and 0.975 quantiles)						
Frailty index, mean*	0.11 (0.04, 0.24)	0.12 (0.06, 0.19)	0.12 (0.07, 0.18)	0.12 (0.08, 0.16)	0.12 (0.09, 0.15)	0.12 (0.10, 0.14)
Prevalence of frailty (FI > 0.28), median	0.11 (0.02, 0.33)	0.10 (0.02, 0.27)	0.06 (0.02, 0.16)	0.04 (0.02, 0.10)	0.04 (0.02, 0.07)	0.03 (0.02, 0.06)
Frailty index cut-off identifying 15% as having frailty	0.20 (0.13, 0.40)	0.20 (0.11, 0.30)	0.20 (0.13, 0.29)	0.20 (0.16, 0.25)	0.20 (0.16, 0.23)	0.19 (0.16, 0.22)
Predictive uses of the frailty index (0.025 and 0.975 quantiles)						
Odds ratio for mortality over the follow-up, per 0.1 FI increase	1.24 (1.02, 1.44)	1.44 (1.18, 1.69)	1.60 (1.31, 1.88)	1.87 (1.58, 2.18)	2.04 (1.79, 2.34)	2.19 (1.97, 2.47)
Predicted probability of mortality over the follow-up for FI = 0.28	0.08 (0.06, 0.12)	0.10 (0.07, 0.15)	0.11 (0.08, 0.17)	0.14 (0.10, 0.19)	0.15 (0.12, 0.21)	0.17 (0.14, 0.21)
Frailty index cut-off predicting 10% mortality over the follow-up	0.38 (0.18, 1.23)	0.28 (0.18, 0.46)	0.25 (0.17, 0.35)	0.22 (0.17, 0.28)	0.21 (0.18, 0.25)	0.20 (0.18, 0.23)
Measurement statistics (95% confidence interval)						
Intraclass correlation coefficient for agreement	0.19 (0.18, 0.19)	0.34 (0.33, 0.34)	0.45 (0.44, 0.46)	0.63 (0.62, 0.63)	0.75 (0.75, 0.76)	0.84 (0.84, 0.85)
Standard error of measurement for agreement	0.13 (0.13, 0.13)	0.09 (0.09, 0.09)	0.07 (0.07, 0.07)	0.05 (0.05, 0.05)	0.04 (0.04, 0.04)	0.03 (0.03, 0.03)

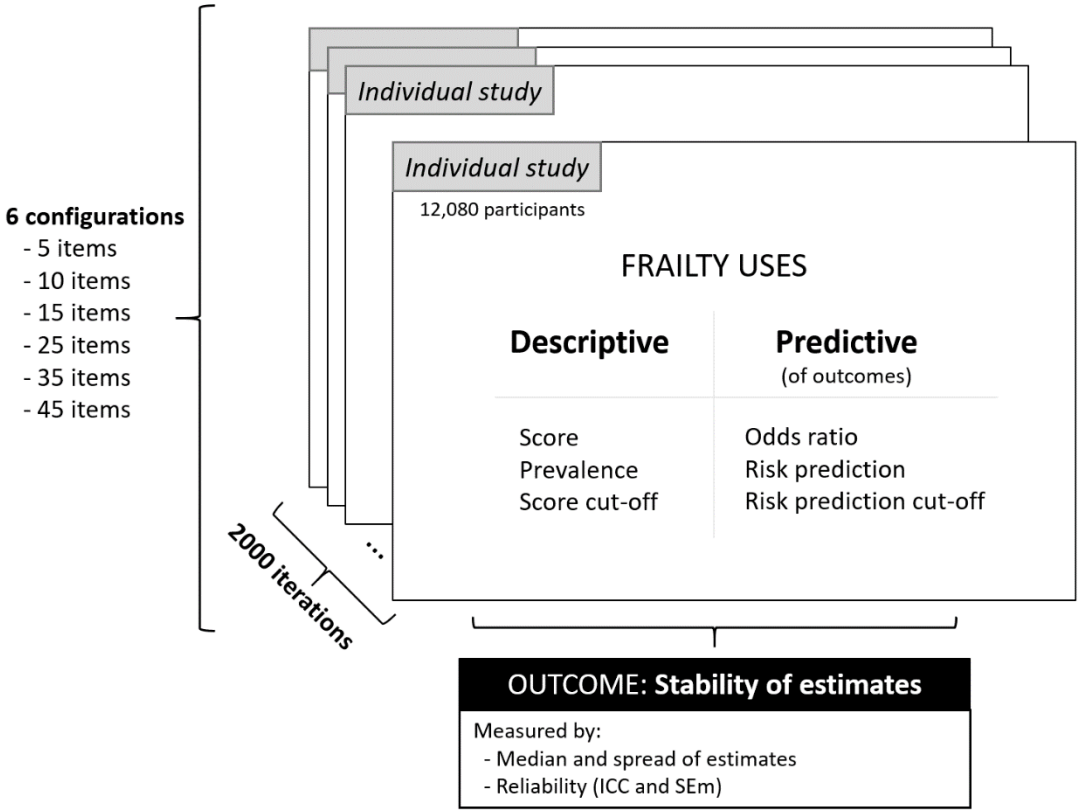
Notes. 2000 iterations (simulated studies) were performed for each configuration. FI = frailty index; *The median, 0.025 and 0.975 quantiles are reported for the distribution of mean frailty index scores in each of 2000 simulations.

Table 5.3. Recommendations for clinical interpretation and usage, and research using frailty indices

CLINICAL INTERPRETATION AND USAGE
Frailty indices and categorical thresholds should be applied with considerable caution due to potentially low reliability. Frailty indices are not interchangeable. Clinical interpretation and application should carefully consider the specific composition of items and the thresholds used in each definition. For studies using frailty indices constructed using a low number of items, the interpretation of results should be restricted to the specific frailty index and deficits considered. Frailty indices should include at least 30 items as has been previously recommended by Searle et al.[42], and ideally 45 items and more. Measures using fewer items may be used but should not be considered comparable to other frailty indices. When interpreting results from frailty indices as a whole, regression dilution should be considered whereby frailty indices comprising a fewer number of items will bias the true magnitude of association toward the null.
REPORTING AND RESEARCH
Frailty indices and their constitutive items should be characterized in greater detail: specific items, item cut-offs, cut-offs for frailty status should be systematically reported. Consider anchoring current and future frailty indices by describing their distribution in a freely available standard population of older adults, thus allowing comparisons. Consider using Item Response Theory methods[38] to further characterize the most frequently used deficits and cut-offs, to compare of frailty indices, and potentially to devise a standard blueprint for frailty items to be included.

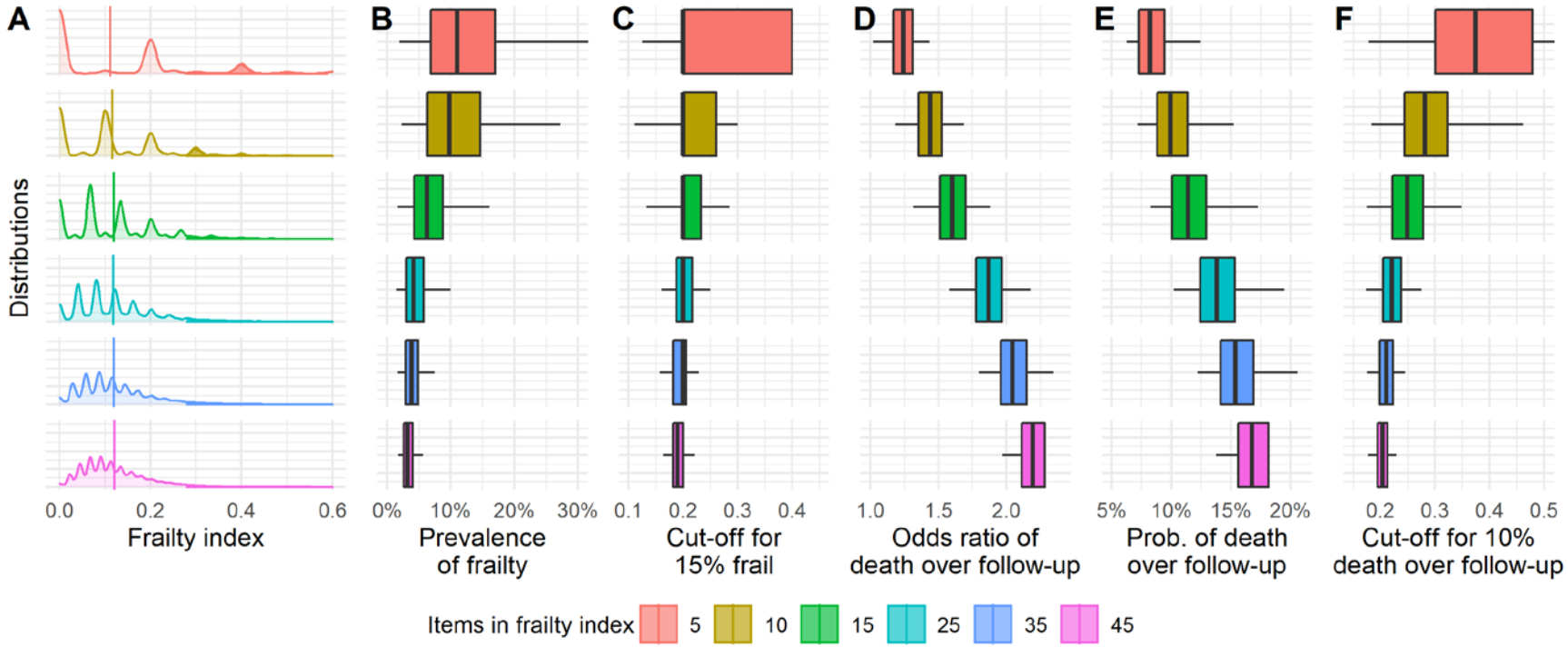
5.12 FIGURES

Figure 5.1. Methodological framework for appraising the clinical measurement properties of various configurations of the frailty index



Notes. ICC = intraclass correlation coefficient, SEm = standard error of measurement

Figure 5.2. Summary of simulation results for six configurations of the frailty index: frailty score, prevalence, frailty cut-offs, odds ratio, and mortality prediction (six configurations x 2000 iterations x 12,080 participants)



Notes. 2A Non-smooth frailty index (FI) densities: the overall area under each curve represents the full distribution of FI values, the height is the relative proportion of FI values; the vertical lines indicate the median and the shaded areas represent those with frailty (prevalence) defined as FI > 0.28. **2B-F** Boxplots: middle line in box indicates the median, box indicates the 25th and 75th percentiles, and whiskers indicate the 2.5th and 97.5th percentiles. **2E** Probability of death over follow-up for FI = 0.28 (prevalence cut-off).

5.13 SUPPLEMENTAL MATERIAL

Supplementary Table 5.1. Deficits, cut-offs, and missing data of deficits used to derive frailty indices and mapping to the original 70 times from the Canadian Study on Health and Aging

Domains	CLSA items	Cut-off if continuous or multiple-category variable	Missingness (proportion)	CSHA items
Physical function measures and mobility (d = 14)	Able to walk	-	0.00	Impaired mobility
	Osteoarthritis of hand, hip, or knee	-	0.00	Musculoskeletal problems
	Trouble rising from chair	-	0.01	Bradykinesia of the limbs
	Max grip strength attained on all trials	Lowest quintile by sex = 1	0.09	Poor muscle tone in limbs
	Pain or paralysis in hands or wrists	-	0.00	Poor limb coordination
	Average time for 1 chair rise (in seconds)	Lowest quintile by sex = 1	0.07	Poor coordination, trunk
	Balance poor	-	0.01	Poor standing posture
	Total time required to complete 4m walk (in seconds)	Lowest quintile by sex = 1	0.02	Irregular gait pattern
	Best attained time - Standing Balance	Lowest quintile by sex = 1	0.09	Impaired vibration
	Arms or legs shake	-	0.00	Tremor at rest
	Shuffle feet	-	0.01	Postural tremor
	Trouble buttoning buttons	-	0.00	Intention tremor
	Parkinsonism or Parkinson's Disease	-	0.00	History of Parkinson's disease
	Falls last 12 months	-	0.03	Falls
Physiological function (d = 1)	Forced expiratory volume after 1 second – Trial1	Lowest quintile by sex = 1	0.30	Respiratory problems
Function and disability (d = 6)	Able to dress	-	0.00	Problems getting dressed
	Able to take bath	-	0.00	Problems with bathing
	Able to take care of appearance	-	0.00	Problems with carrying out personal grooming
	Trouble to get in time to bathroom	-	0.00	Toileting problems
	Able to prepare meals	-	0.00	Problems cooking
	Able to go shopping	-	0.00	Problems going out alone
Activity-exhaustion-energy (d = 1)	CES-D 10 scale: Frequency feel could not 'get going'	> 2 days/week = 1	0.01	Tiredness all the time
Cognition (d = 11)	Memory problem	-	0.00	Memory changes
	REYII - Number of words (or variants) correctly recalled in 90 seconds - Delayed Recall	Lowest quintile = 1	0.04	Short-term memory impairment

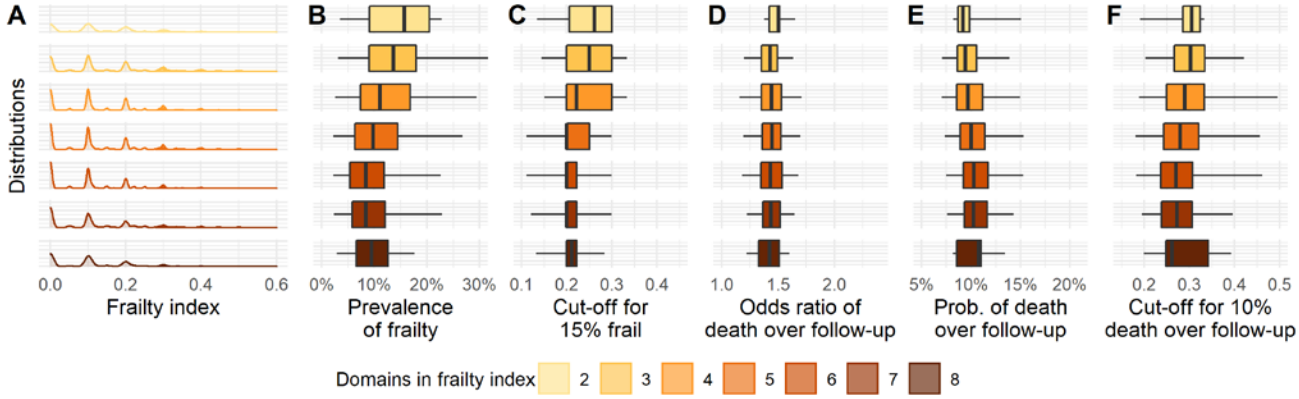
	Three target actions	Correct 3 = 0 Correct 2 = 0.5 Correct ≤ 1 = 1	0.00	Long-term memory impairment
	Number of different animals recited in 60 seconds	Lowest quintile = 1	0.03	Onset of cognitive symptoms
	Able to count from 1 to 20	-	0.02	Clouding or delirium
	CES-D 10 scale: Frequency easily bothered	5-7 days/week = 1 3-4 days/week = 0.5 ≤ 2 days/week = 0	0.01	Paranoid features
	Traumatic brain injury	-	0.00	History relevant to cognitive impairment or loss
	Stroop[1.2] - Number of errors – Interference	Lowest quintile = 1	0.02	Family history relevant to cognitive impairment or loss
	Number of words (or variants) correctly recalled in 90 seconds - Immediate Recall	Lowest quintile = 1	0.04	Family history of degenerative disease
	Able to feed	-	0.00	Presence of snout reflex
	Able to get out of bed	-	0.00	Presence of palmomental reflex
Medical conditions and symptoms (excluding mobility and psychiatric) (d = 26)	Usually free of pain and discomfort	-	0.03	Head and neck problems
	Back pain past one month	-	0.01	Poor muscle tone in neck
	Face less expressive	-	0.05	Bradykinesia, facial
	Frequency of incontinence	-	0.00	Urinary incontinence
	Bowel incontinence	-	0.00	Rectal problems
	Sleep quality	Very dissatisfied = 1 Dissatisfied = 0.5 Very satisfied, satisfied, neutral = 0	0.00	Sleep changes
	Frequency restless or fidgety	All of the time, most of the time = 1 Some of the time, a little of the time, none of the time = 0	0.03	Restlessness
	Epilepsy	-	0.00	Seizures, partial complex
	Ever had cataracts	-	0.02	Seizures, generalized
	Sudden loss of vision in one eye	-	0.01	Syncope or blackouts
	Migraine headaches	-	0.00	Headaches
	Experienced a ministroke or TIA	-	0.01	Cerebrovascular problems
	Stroke or CVA	-	0.01	History of stroke
	Diabetes, borderline diabetes, or high blood sugar	-	0.00	History of diabetes mellitus
	High blood pressure or hypertension	-	0.01	Arterial hypertension

	Peripheral vascular disease or poor circulation in limbs	-	0.01	Peripheral pulses
	Unstable heart condition within last 3 months	-	0.00	Cardiac problems
	Heart attack or myocardial infarction	-	0.01	Myocardial infarction
	Coronary artery bypass surgery	-	0.00	Arrhythmia
	Heart disease (including congestive heart failure, or CHF)	-	0.01	Congestive heart failure
	Emphysema, chronic bronchitis, COPD, or chronic changes in lungs due to smoking	-	0.01	Lung problems
	Under-active thyroid gland	-	0.02	History of thyroid disease
	Over-active thyroid gland	-	0.02	Thyroid problems
	Osteoporosis	-	0.01	Skin problems
	Cancer	-	0.00	Malignant disease
	Breast cancer	-	0.00	Breast problem
Mood and psychiatric conditions and symptoms (d = 4)	Mood disorder	-	0.00	Mood problems
	CES-D 10 scale: Frequency feel depressed	5-7 days/week = 1 3-4 days/week = 0.5 ≤ 2 days/week = 0	0.01	Feeling sad, blue, depressed
	Anxiety disorder	-	0.00	History of depressed mood
	Center for Epidemiological Studies Short Depression Scale (CES-D 10) score	> 15 = 1 10-14 = 0.5 0-9 = 0	0.01	Depression (clinical impression)
Nutrition, weight (d = 4)	Bowel disorder	-	0.01	Gastrointestinal problems
	Intestinal or stomach ulcers	-	0.01	Abdominal problems
	Cough, choke pain when swallowing food	Always, often, sometimes = 1 Rarely, never = 0	0.03	Sucking problems
	Low appetite	Poor = 1 Fair = 0.5 Good, very good = 0	0.03	Bulk problems
Subjective health (d = 2)	Self-rated general health	Good, fair, poor = 1 Very good = 0.5 Excellent = 0	0.00	Changes in everyday activities
	Perceived mental health	Poor = 1 Fair = 0.5 Good, very good, excellent = 0	0.00	Changes in general mental functioning
Health usage (d = 1)	Seen in an emergency department in past 12 months	-	0.03	Other medical history

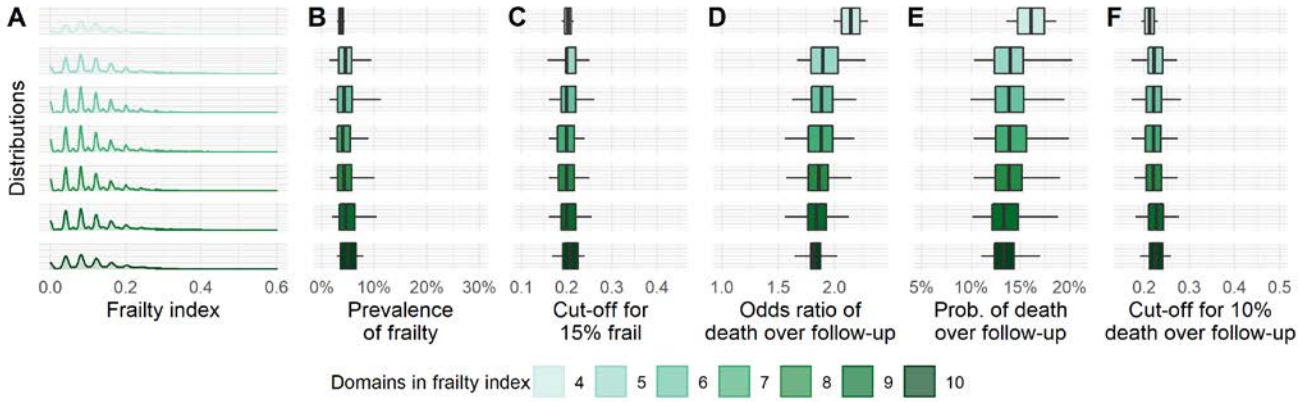
Notes. CLSA: Canadian Longitudinal Study on Aging, CSHA: Canadian Study of Health and Aging

Supplementary Figure 5.1. Summary of simulation results for 3 configurations of the frailty index (10, 25, and 45 items), by number of domains included: frailty score, prevalence, frailty cut-offs, odds ratio, and mortality prediction

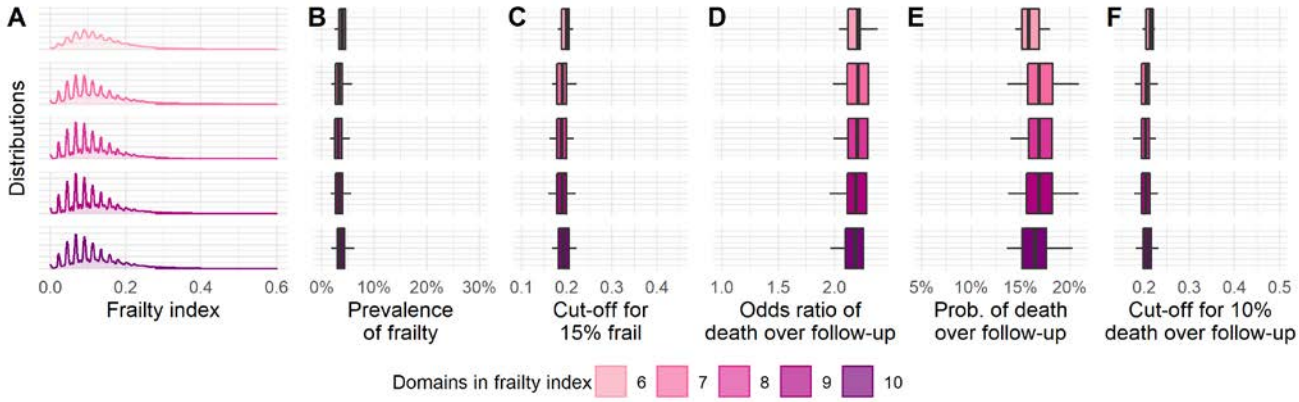
10-item frailty indices



25-item frailty indices

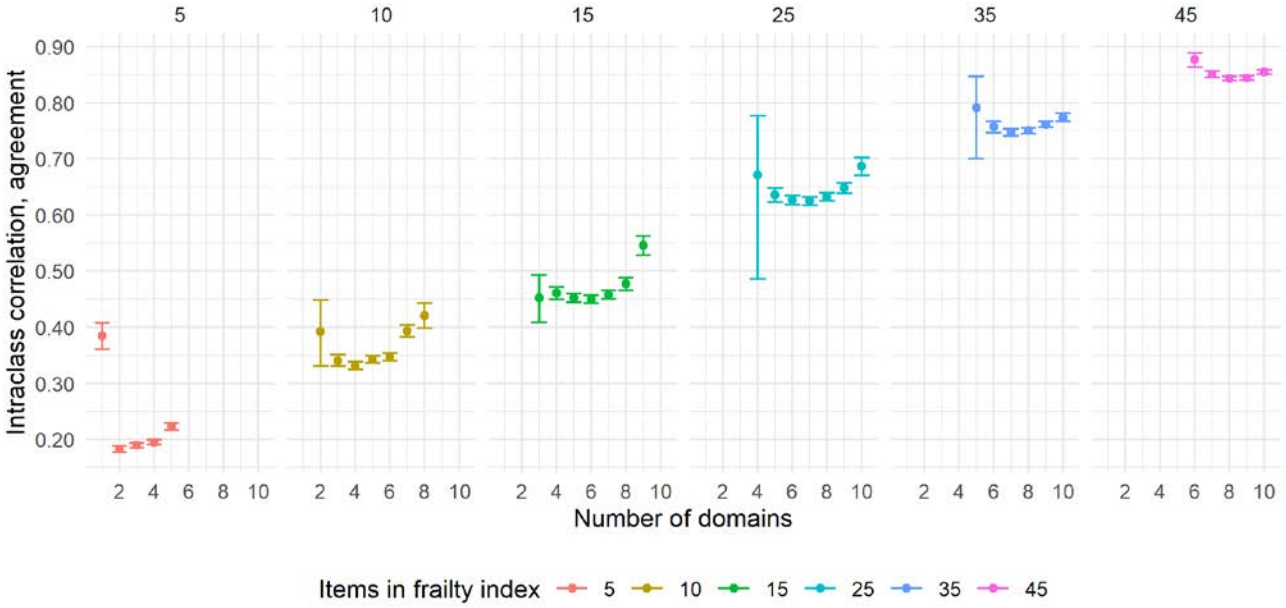


45-item frailty indices



Notes. A Non-smooth frailty index (FI) densities: the overall area under each curve represents the full distribution of FI values, the height is the relative proportion of FI values. **B-F** Boxplots: middle line in box indicates the median, box indicates the 25th and 75th percentiles, and whiskers indicate the 2.5th and 97.5th percentiles. **E** Probability of death over follow-up for FI = 0.28 (prevalence cut-off).

Supplementary Figure 5.2. Intraclass correlation coefficients of frailty indices by number of items and number of domains



Notes. Intraclass correlation coefficients for agreement (ICC_A) increase with the greater number of items in frailty indices. The number of items accounts for more variation in ICC_A than the specific number of domains included in each configuration. The high ICC_A reported for the 5-item frailty indices including only one domain can be explained by the greater item correlation in frailty indices measuring a single domain (e.g., physical function measures and mobility, cognition, or medical conditions and symptoms). Of note, although these 5-item frailty indices measure deficit accumulation, they do not capture multiple dimensions, an inherent characteristic of frailty.

Additional references for supplemental information

1. Troyer AK, Leach L, Strauss E. Aging and Response Inhibition: Normative Data for the Victoria Stroop Test. *Aging, Neuropsychol Cogn*. 2006;13(1):20-35. doi:10.1080/138255890968187

2. Bayard S, Erkes J, Moroni C, Roussillon) the C des PC spécialisés en N du LR (CPCN L. Victoria Stroop Test: Normative Data in a Sample Group of Older People and the Study of Their Clinical Applications in the Assessment of Inhibition in Alzheimer's Disease. *Arch Clin Neuropsychol*. 2011;26(7):653-661. doi:10.1093/arclin/acr053

Chapter 6. Manuscript 3: Appraising clinical applicability of studies: mapping and synthesis of current frameworks, and proposal of the FrACAS framework and VICORT Checklist

6.1 PREFACE

Older adults are heterogeneous compared to younger adults on select health characteristics which can be quantified using age-related constructs such as frailty. Even when there is a basis for categorizing individuals, this categorization should be reliable. Even if this categorization is substantiated and reliable, the work of that nosological construct is not yet complete from a *clinical imperative* perspective. Not all constructs are required to fulfill the clinical imperative of materially making a difference on health outcomes; however, one that aims for clinical implementation should. In Manuscript 3, I first used a systematic literature search to identify, map, and synthesize frameworks and criteria that appraise *applicability* of studies and study findings. Applicability holds multiple understandings in the literature, and its broad meaning may not determine whether a construct under study can alter or improve health outcomes. Using the findings from the literature review and mapping, supplemented with themes from contemporary debates in epidemiology and clinical medicine, I proposed a novel framework to specifically appraise the clinical applicability of studies. Manuscript 3 sought to identify the specific criteria that determine whether studies can be relevant (and applicable) to clinical practice. It proposes an appraisal framework with six underlying criteria.

***

This manuscript has been submitted to BMC Medical Research Methodology.

6.2 TITLE PAGE

Appraising clinical applicability of studies: mapping and synthesis of current frameworks, and proposal of the FrACAS framework and VICORT Checklist

Short title

Appraising clinical applicability of studies

Authors and affiliations

Quoc Dinh Nguyen, MD MA MPH,[a,b,c] Erica M. Moodie, PhD,[c] Philippe Desmarais, MD MHSc,[a,b] Robert Goulden, MBBS MSc MA,[c,d] Marie-France Forget, MD MSc,[a] Eric Peters, MD,[e] Sahar Saeed, PhD,[f,g] Mark R. Keezer, MDCM PhD,[b,h] Christina Wolfson, PhD[c,d,i]

a. Division of Geriatrics, Centre hospitalier de l'Université de Montréal, Canada;
b. Centre de recherche du Centre hospitalier de l'Université de Montréal, Canada;
c. Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Canada;
d. Department of Medicine, McGill University, Canada;
e. Department of Anesthesia, Centre hospitalier universitaire Sainte-Justine, Canada;
f. Canadian Blood Services, Ottawa, Canada
g. Department of Infectious Disease, Washington University in St. Louis, St. Louis, USA
h. Departments of Neurosciences & Social and Preventative Medicine, Université de Montréal, Canada
i. Neuroepidemiology Research Unit, Research Institute of the McGill University Health Centre, Canada.

Corresponding author

Quoc Dinh Nguyen
1000, Saint-Denis, Montreal, Quebec, Canada H2X0C1
quoc.nguyen@mail.mcgill.ca

Word counts

Abstract: 245
Manuscript: 3658
Tables: 2
Figures: 3
References: 80

6.3. ABSTRACT

Background

Not all research findings are translated to clinical practice. Reasons for lack of applicability are varied, and multiple frameworks and criteria exist to appraise the general applicability of epidemiological and clinical research. In this two-part study, we first identify, map, and synthesize frameworks and criteria; we develop a framework to assist clinicians appraise applicability specifically from a clinical perspective.

Methods

We conducted a literature search in PubMed and Embase to identify frameworks appraising applicability of study results. Conceptual thematic analysis was used to synthesize frameworks and criteria. We proposed an applicability appraisal framework and six-criteria checklist by integrating four inputs: contemporary debates in epidemiology, brainstorming and discussions, findings from the literature search and synthesis, and iterative pilot-testing.

Results

Of the 4622 references retrieved, we identified 26 unique frameworks featuring 21 criteria. Frameworks and criteria varied by scope and level of aggregation of the evidence appraised, target user, and specific area of applicability (internal validity, clinical applicability, external validity, and system applicability). Our proposed framework classifies studies in three domains (research, practice informing, and practice changing) by examining six criteria sequentially: *V*alidity, *I*ndication-informativeness, *C*linical relevance, *O*riginality, *R*isk-benefit comprehensiveness, and *T*ransposability (VICORT checklist).

Conclusions

Existing frameworks to applicability vary by scope, target user, and area of applicability. We introduce our concise *Fr*amework *A*ppraising the *C*linical *A*pplicability of *S*tudies (FrACAS) which specifically assessed applicability from a clinical perspective. Our framework can be used as a tool for the design, appraisal, and interpretation of epidemiological and clinical studies.

Keywords

Quality assessment, external validity, generalizability, impact, evidence-based practice

6.4. INTRODUCTION

Not all health research findings are translated into clinical or public health interventions.[1] Many reasons for lack of implementation can relate to research quality and validity.[2–5] Excellent frameworks have been developed to assess the quality of epidemiological and clinical research by predominantly assessing the internal validity of research findings (e.g., confounding, selection and measurement biases).[6–9] What determines high quality and validity research may not, however, directly determine what is most impactful.[10] The appraisal of *applicability*, whether study results can impact practice, demands an expanded set of considerations. The cumulative nature of evidence and of the strength of evidence is the focus of many important frameworks, most notably GRADE (Grading of Recommendations, Assessment, Development and Evaluations)[11] used to synthesize evidence and formulate clinical recommendations. External validity is another critical focus when applying study results to specific practice and population contexts (generalizability and transportability).[12–15] Implementation science and economic considerations also factor in the practical application of research.[16–19]

Although current frameworks cumulatively cover many important facets of applicability, the specific criteria to assess applicability may vary by the type of research and evidence, and by the stakeholders involved: researchers, clinicians, decision-makers and policy-makers. *Clinical* applicability can be defined as the potential of study findings to inform or directly alter current clinical practice at the individual level. Due to their wide scope, it is unclear whether existing frameworks can concisely assist clinicians in differentiating between studies that change practice, inform practice, or are not clinically applicable. As clinicians must evaluate an ever-expanding research output, there is a need to better identify criteria that may be used to gauge applicability, in particular clinical applicability.

In this two-part study, we conducted a broad literature review to identify, map, and synthesize existing frameworks and criteria pertaining to the applicability of studies. Drawing from this review, current concepts and debates in epidemiology[20–23] and clinical research,[24,25] and iterative discussions and testing, we developed a concise tool to classify and improve the applicability of studies, with an emphasis on the clinical perspective. FrACAS, our proposed Framework to Appraise the Clinical Applicability of Studies and its checklist (VICORT) are introduced and discussed.

6.5. METHODS

6.5.1. Search, thematic mapping, and synthesis of available frameworks

We searched PubMed and EMBASE (Ovid) databases since their inception for articles reporting on frameworks appraising the general "applicability" of research findings on November 12, 2020. The eligibility criteria were articles (i) featuring a unique tool, instrument, checklist, or framework (ii) focused on the applicability to practice or (iii) health research evidence, and (iv) published in English. We excluded articles that solely featured a review of frameworks, the application of an existing framework, or were restricted to a specific condition or discipline. Due to the potential multiple understandings of "applicability," we used combinations of keywords in titles and abstracts to maximize the comprehensiveness of article selection as previously done by others on the topic of applicability;[12,13] the full search strategy is detailed in the Supplementary Methods. Duplicates were removed, titles and abstracts were screened independently by two authors (PD and QDN). We supplemented remaining articles with references in reviews and retrieved articles. Articles were assessed in full to identify unique frameworks. PD and QDN performed conceptual thematic analysis[26] using preliminary themes that were refined iteratively to map the frameworks

and to synthesize criteria of applicability by stakeholders. Disagreements were resolved by consensus.

6.5.2. Development of framework for clinical applicability

As illustrated in Supplementary Figure 6.1, we developed our framework by integrating four major inputs: contemporary debates in epidemiology and clinical research, brainstorming and discussion meetings, comparison with existing frameworks for appraisal of clinical applicability, and pilot application of iterative versions of our frameworks in a scoping review on clinical frailty (forthcoming). Brainstorming and discussion meetings involved contributions from clinicians, researchers, and methodologists with expertise in multiple substantive domains of clinical practice and research (intensive care, pediatrics, neurology, internal, emergency, and geriatric medicine), as well as epidemiology, biostatistics, qualitative, and translational research. Formal Delphi methodology was not employed; preliminary versions of the framework were iteratively tested and refined to reach the final consensus framework.

6.6. RESULTS

6.6.1. Analysis, mapping, and synthesis of frameworks for applicability

We identified 4622 references, of which 1324 were duplicates and 3265 were excluded following the screening of titles and abstracts, leaving 33 for assessment. Thirty additional references were identified in reviews and references from retrieved articles; we assessed 63 full-text articles and included 26 unique frameworks. Supplementary Figure 6.2 presents the flowchart for article selection.

6.6.1.1. Description and analysis of frameworks

Table 6.1 presents the 26 frameworks and their predominant focus.[6,7,11,14,15,19,27–53] Frameworks were published between 1999 and 2021 in epidemiological, clinical, public health, policy, and

decision-making journals. Although we only included frameworks related to applicability, the focus varied widely from the quality of clinical practice guidelines (CPG, AGREE I-II[50,51]), quality and strength of recommendations (GRADE),[11] use of evidence to inform health decisions (GRADE EtD),[15] applicability of prediction model studies (PROBAST),[37] applicability of randomized trials (PRECIS)[35] and health technology assessments (HTA).[41,49] Due to distinct purpose and focus in appraising applicability, the complexity of frameworks and the number, nature, and level of criteria detail within frameworks also varied. Some frameworks featured a simple list of key criteria[44,47] whereas others elaborated on a full system of domains, criteria, and appraisal processes (e.g., RE-AIM,[19,38] GRADE,[11] PRECIS,[35] RoB2,[7] RoBINS-I,[54] Atkins et al.[42]); some adapted to specific concepts and disciplines (GRADE EtD).[15,28–32] After comparative analysis of frameworks, we identified three dimensions explaining the variability which we used to map the frameworks and criteria:

- The primary intended target user or stakeholders (researchers, clinicians, and decision-makers);
- The evidence type appraised and its level of aggregation, from fundamental research to CPG;
- The areas of applicability: internal validity, clinical applicability for individual patients, external validity, and applicability at the system level.

Although the categories within these dimensions are not mutually exclusive, they allow the mapping and synthesis of the multiple purposes and understandings of applicability, as illustrated in Figures 6.1 and 6.2.

6.6.1.2. Mapping of frameworks and synthesis of criteria

Figure 6.1 maps the 26 frameworks according to the evidence type appraised and the primary intended target user. For most frameworks, the scope of the evidence appraised was directed at a single level of aggregation (e.g., prediction studies,[33,37] trials,[7,14,35,42,44,45] CPG [48,50,52,55]); a few frameworks bridged evidence types such as the GRADE[11] framework which examines findings from case-control and cohort studies to systematic reviews. Most frameworks were intended for multiple stakeholders (researchers, clinicians, decision-makers), but none encompassed all three. There was a qualitative association between the level of aggregation of evidence and the primary intended users: as the frameworks appraised increasingly aggregated evidence (e.g., HTA or CPG) the target users tended toward decision-makers, whereas frameworks pertaining to prediction and observational studies were more focused on researchers, with in the middle, frameworks on trials focused mostly on clinicians.

Figure 6.2 summarizes the criteria extracted from the frameworks. Across all frameworks, 21 criteria were synthesized and qualitatively mapped to evidence type appraised and the applicability areas. Although there was overlap of areas of applicability, 7 criteria fell under internal validity (i.e., risk of bias, confounding, reporting bias, dose-response gradient, precision, directness, consistency of results, and comparison intervention); 6 criteria under applicability at the system level (i.e., acceptability and feasibility, sustainability, cost and cost-effectiveness, scope of practice and actions, equity and ethics, monitoring/audit and support tools). In between, clinical applicability at the individual level directly encompassed 5 criteria (i.e., comparison intervention, intervention characteristics, magnitude and trade-offs of harms and benefits, relevance of outcomes, strength/level of evidence); and external validity considered 3 critical criteria (values, beliefs, preferences priority; context and resources for application; representativeness of patients

and populations). The latter two criteria along with relevant outcomes were the most frequently featured criteria across frameworks. There was a qualitative association between criteria in frameworks about higher level of aggregation of evidence and applicability at the system level. Existing frameworks on clinical applicability span multiple target users, evidence types, and areas of applicability. Applicability holds different meanings whether one is a researcher, clinician, or decision-maker, and is ascertained using different set of criteria depending on the type of evidence and whether internal validity, clinical applicability, external validity, or system applicability is emphasized. Our proposed framework focuses on the clinical perspective and aims to assist *clinicians* when evaluating *all types* of *primary study results* (from fundamental research to RCT and trials) to determine whether and how these apply to *clinical practice*.

6.6.2. Proposed framework: the Framework to Appraise the Clinical Applicability of Studies (FrACAS) and VICORT checklist

6.6.2.1. Operational definition and classification of "clinical applicability:" the FrACAS framework

FrACAS uses an operational definition of clinical applicability that classifies a study according to the following questions: "are these research results valid?", "can these results *inform* [my] practice?", or "do these results *change* [my] current practice?". As shown in Figure 6.3, studies are classified in one of three evidence domains: research, practice-informing, or practice-changing domains, based on six criteria that examine study design elements and related data sources.

6.6.2.2. Criteria for appraisal and classification in FrACAS: the VICORT checklist

The six criteria that determine study classification in FrACAS are: *V*alidity, *I*ndication-informativeness, *C*linical relevance, *O*riginality, *R*isk-benefit comprehensiveness, and *T*ransposability (VICORT checklist). Study findings are considered progressively more

informative and practice changing as they sequentially meet these criteria. Table 6.2 presents each criterion's definition and comparisons with criteria synthesized in the review.

6.6.2.2.1. Validity

Validity is the criterion most discussed, established, and assessed by researchers and clinicians .[2,3] Internal validity is a necessary criterion for study findings to be considered research evidence. As our review shows, most quality assessment tools, including the Cochrane Risk-of-Bias tool (RoB 2)[7] and the Risk Of Bias In Non-randomised Studies of Intervention (ROBINS-I),[6] focus on the validity of methods (randomization, blinding, and missing data; confounding, information, and endogenous selection bias). The importance of validity in general applicability of study results is highlighted by the 7 validity-related criteria shown in Table 6.2. Although validity is a prerequisite, it is not sufficient for clinical applicability.

6.6.2.2.2. Indication-informativeness

Validity ensures that estimates are unbiased. Indication-informativeness ensures that these estimates are applicable in clinical practice. Study findings produce estimates, but not all estimates can lead to action in clinical practice. To do so, the study should produce results that inform a *clinical indication*, i.e., an *intervention* in a *specific population*. An indication entails the identification of *what* clinicians should *do* and *which population* would benefit from this being done. To inform a clinical indication, a study must include a well-defined intervention whose effect is identifiable in the results (i.e., identifiability). The ability to identify and to promise the future effects of an intervention under consideration is the key criterion to achieve indication-informativeness and move from the research domain to the clinical practice domain.

Only some study designs fulfill this criterion. Firstly, randomized control trials (RCT) where an intervention is evaluated in an eligible/target population. Secondly, observational studies

of an exposure for which there exists an intervention (or where one is envisioned) to remove or modify the exposure of interest.[56] If validity is ensured, the effect of the intervention can be identified and generally assumed to approximate the effect of the exposure (e.g., smoking cessation and smoking). The existence (or lack thereof) of an exposure-removing intervention is the core of the indication-informativeness criterion. HIV, smoking, atherosclerosis, frailty, and age are exposures with decreasing levels indication-informativeness since eliminating each is increasingly challenging. Third, observational studies can also inform a clinical indication by descriptively reporting absolute outcomes of an *already/otherwise-indicated* intervention in a *specific* population of interest. For example, reporting the absolute mortality following heart surgery indicated for coronary artery disease, in patients with frailty, informs this indication by allowing the counterfactual contrast between undergoing an intervention and the natural history when forgoing the intervention, *in those with frailty.* Of note in this scenario, the well-defined intervention is not indicated on the basis of frailty. Following these three study designs, exposures can form the basis of an indication (i.e., inform an intervention or specific population) only when they are used in a study as a selection criterion, predictor, mediator, or effect modifier, not when used as a confounder or outcome.

Indication-informativeness does not currently feature explicitly in any identified frameworks. However, it is strongly related to the widely debated requirement of well-defined interventions in epidemiology.[20,57–59] Our framework contextualizes the presence of the well-defined intervention/consistency assumption[23,60] as a requirement for evidence that is clinically informative and applicable, not for epidemiological evidence itself.[61]

6.6.2.2.3. Clinical relevance

Epidemiological research spans a broad range of outcome types including basic science mechanisms, intermediate outcomes, and patient-centered outcomes.[25] Clinical relevance requires that study outcomes be directly relevant and informative to practice. The precise delimitation of what outcomes are informative to practice varies.[25] It may be easy to restrict measures of heart stem cell transplantation survival to being clinically non-informative, but cholesterol levels, coronary calcium scores, atherosclerotic cardiovascular disease hospitalization, mortality, and health-related quality of life (HRQoL) all have some clinically relevant information. Achieving full clinical relevance benefits from incorporating patient-centered outcomes, of which mortality and HRQoL are examples. Ignoring outcomes that are patient-centered has led to increased numbers of studies using surrogate outcomes with unclear patient benefit and potential overdiagnoses.[24,62] Clinical relevance in FrACAS is related to the directness[11,63] and relevance of outcomes criteria identified in our review.

6.6.2.2.4. Originality: clinical significance and novelty

The originality criterion comprises significance and novelty. Under our framework, significance centers on demonstrating a clinically meaningful magnitude of effect (effect size), not only statistical significance.[64] Even if results are clinically meaningful, they can only *alter* current practice if they are novel compared to the current evidence base and standard practice, as shown in Figure 6.3. Appraising novelty requires contrasting study results with a careful examination of the cumulative substantive evidence (e.g., reviews, practice guidelines) and current practices. Appraisal is thus practice-setting dependent. The novelty of a study involves changing an intervention-population coupling: this requires altering (i.e., adding or removing) an intervention in a specific population or, conversely, modifying a specific population as eligible for an

intervention. For example, finding that exercise benefits older adults *with frailty* may not be novel since exercise is *already* recommended to older adults *in general*. The difference between statistical and clinical significance (magnitude of benefits) has been highlighted in frameworks,[11,14,15,34,40,41,52] but the importance of the novelty of findings to alter practice has not. The lack of novelty may explain why some prediction studies do not alter practice: if all modifiable predictive exposures are already addressed in standard care, then no new indication can be identified.

6.6.2.2.5. Risk-benefit comprehensiveness

Will altering an indication in current practice prove *comprehensively* beneficial to patients? Two sides must be examined: first, the intervention and displaced alternatives and, second, their summary net effect on overall outcomes.[65] Comparing a drug to placebo will not displace the same alternatives as comparing a drug with another active agent; if the study outcome is condition-specific at the expense of remaining patient-centered, important complications or outcomes may be overlooked that would outweigh the observed benefit. The withdrawal of the nonsteroidal anti-inflammatory drug rofecoxib due to unanticipated cardiovascular events is one example of the importance of comprehensively considering risks and benefits.[66] The risk-benefit comprehensiveness criteria emphasizes the necessity of examining explicitly and comprehensively the magnitude and trade-offs of harms and benefits criterion identified in available frameworks.[11,14,15,34,40,41,52] The correct calculation of comprehensive health outcomes to estimate net-benefit requires that outcomes be integrated on the absolute scale rather than on the relative scale.[67]

6.6.2.2.6. Transposability

Appraising transposability involves taking all elements of study design, including the broader context of the study, and applying them to a specific practice setting. Epidemiologists and clinicians readily consider the external validity rubrics of generalizability and transportability.[22,68,69] Our transposability criterion has a wider scope. In addition to considering the population and effect modifiers (effectiveness),[22] transposability includes all other facets of implementing the intervention in a given practice setting, e.g., acceptability and feasibility, cost-effectiveness, ethics, and sustainability.[15,19,40,42,47] These will vary by practice context: resource settings, income levels, health care systems and payers, preferences priority, etc.[15,18,40,46,70] As these additional questions enter into the realm of implementation science and economic evaluation, they may be beyond the direct purview of epidemiological research and are not exhaustively detailed in FrACAS. Under our framework, full transposability is optional to reaching the practice-changing evidence domain. But because all of its facets are required when applying study results to a practice setting, transposability should be acknowledged when appraising clinical applicability.

6.7. DISCUSSION

We identified 26 unique frameworks that appraise applicability of studies varying according to the evidence type assessed and the intended target user. Within these frameworks we synthesized 21 criteria focused on four facets of applicability (internal validity, clinical applicability at the individual level, external validity, and applicability at the population or system level). Our mapping of frameworks can help researchers, clinicians, and decision-makers select the most suitable framework depending on the appraisal question and context; selected framework may be further customized by including other synthesized criteria.

We propose a framework aiming to assist clinicians in the appraisal of clinical applicability. FrACAS shares many criteria with existing more structured and widely adopted frameworks. We believe that FrACAS is complementary to the established frameworks for four reasons. First, our framework creates three practical and operational domains of clinical applicability that are meaningful from a clinical practice standpoint: research evidence (i.e., does not inform clinical practice directly), practice informing, and practice changing. Rather than having the full body of existing evidence on a topic as the primary area of focus, FrACAS takes each individual study and characterizes its clinical applicability and impact, which is typically how new findings are examined and consumed in daily practice.

Second, to distinguish between levels of evidence domains, FrACAS proposes two additional criteria not explicitly featured in other frameworks: indication-informativeness and originality. Many frameworks emphasize study design to determine clinical applicability and give more weight to RCT and meta-analyses than to cohort and case-control designs.[6] The indication-informativeness criterion makes clear that it is not the study design *per se* that allows a study to inform and alter practice but its ability to validly inform an indication. Many health-improving interventions did not originate from experimental evidence (e.g., smoking cessation). RCT evidence has an easier claim to validity, indication-informativeness, and thus clinical applicability. However, one cannot invalidate causal inference from observational studies, only require more caution.[56] The new criterion of originality is important to differentiate studies between being practice-informing or practice-changing. Determining originality (novelty and significance) is clinically consequential: practice-informing studies can go unnoticed by clinicians without major detriment since they do not alter any indication, but practice-changing studies cannot. The novelty

of study results is often the prime answer to the “so what?” question of clinical applicability, following the “is it credible?” question of internal validity.

Third, our framework and criteria span multiple evidence types and target users, from fundamental research up to trials and, though focused on clinicians, can be relevant to researchers and decision-makers. FrACAS proposes six relatively orthogonal criteria and does not reduce them to one or two dimensions to summarize the strength or certainty of evidence.[71] FrACAS can be used as a checklist to diagnose which study design elements should be addressed for a study to change practice. Clinical translation can and does occur in the absence of one or many criteria, but we believe that careful analysis would reveal that missing criteria are assumed. We believe that the conciseness of our framework and checklist will help clinicians and trainees appraise and discuss study findings in daily practice.

Fourth, our framework emphasizes the highly contextual and potentially subjective nature of appraising clinical applicability. By explicitly describing study design elements and data sources to be examined for each criterion, we show that determining practice-changing status requires the consideration of an increasing number of features. Whereas classifying articles as practice informing can be based on the appraisal of the individual study in question, a practice changing classification requires consideration of the cumulative evidence base, current standard and specific practice setting. Changing practice is an interdisciplinary and concerted effort requiring both methodological and substantive expertise.

6.7.1. Limitations

Although we carried out a robust literature search, extraction, and synthesis process, we did not conduct a formal systematic review. Our review serves primarily as a map to compare frameworks and criteria rather than to examine their relative strengths and weaknesses [12,13,72–74]. The process

of developing a conceptual framework entails some subjectivity and variability; although a formal Delphi method was not employed, we included a wide range of inputs to iterate versions of our framework (current frameworks, debates in epidemiology, multiple stakeholders, and pilot testing). This representativity and the relative overlap with existing frameworks provide face and content validity. Ultimately, the most proper test of validity and usefulness of our framework will be determined in its usage and application in the real world; further refinements may benefit from wider inclusion of patient and institutional stakeholders.

6.8. CONCLUSION

Frameworks appraising applicability can be classified according to the types of evidence assessed, target users, and areas of applicability (internal validity, clinical applicability, external validity, applicability at population/system level). We proposed a concise framework focusing on clinical applicability which uses six criteria to classify studies into three evidence domains: research, practice informing, or practice changing. Our framework can be used as a tool for the design, appraisal, and interpretation of epidemiological and clinical studies to improve their clinical applicability.

6.9. ACKNOWLEDGEMENTS

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

The authors have no funding sources to declare.

Authors' contributions

QDN, PD, EMM, CW, and MRK designed the study. QDN and PD were responsible for data extraction and analysis. QDN, PD, RG, MFF, EP and SS contributed to the qualitative analysis and interpretation of data. All authors contributed significantly to the content, critically reviewed, and provided input during manuscript revision.

6.10. REFERENCES

1. Morris ZS, Wooding S, Grant J. The answer is 17 years, what is the question: Understanding time lags in translational research. *J R Soc Med*. 2011;104(12):510-520. doi:10.1258/jrsm.2011.110180

2. Maclure M, Schneeweiss S. Causation of bias: The episcope. *Epidemiology*. 2001;12:114-122. doi:10.1097/00001648-200101000-00019

3. Greenland S. Validity and bias in epidemiological research. In: Oxford, UK: Oxford University Press; 2011. doi:10.1093/med/9780199218707.003.0037

4. Campbell DT. Factors relevant to the validity of experiments in social settings. *Psychol Bull*. 1957;54(4):297-312. doi:10.1037/h0040950

5. Jüni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. *Br Med J*. 2001;323(7303):42-46. doi:10.1136/bmj.323.7303.42

6. Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*. 2016;355:i4919-i4919. doi:10.1136/bmj.i4919

7. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ*. 2019;366:l4898-l4898. doi:10.1136/bmj.l4898

8. Greenland S, Morgenstern H. Confounding in health research. *Annu Rev Public Heal*. 2001;22:189-212.

9. Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. *Epidemiology*. 2004;15(5):615-625. doi:10.1097/01.ede.0000135174.63482.43

10. Galea S. An argument for a consequentialist epidemiology. *Am J Epidemiol*. 2013;178(8):1185-1191. doi:10.1093/aje/kwt172

11. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008;336(7650):924 LP - 926. doi:10.1136/bmj.39489.470347.AD

12. Burchett H, Umoquit M, Dobrow M. How do we know when research from one setting can be useful in another? A review of external validity, applicability and transferability frameworks. *J Heal Serv Res Policy*. 2011;16(4):238-244. doi:10.1258/jhsrp.2011.010124

13. Burchett HED, Blanchard L, Kneale D, Thomas J. Assessing the applicability of public health intervention evaluations from one setting to another: A methodological study of the usability and usefulness of assessment tools and frameworks. *Heal Res Policy Syst*. 2018;16(1):15-17. doi:10.1186/s12961-018-0364-3

14. Green LW, Glasgow RE. Evaluating the relevance, generalization, and applicability of research: Issues in external validation and translation methodology. *Eval Heal Prof*. 2006;29(1):126-153. doi:10.1177/0163278705284445

15. Alonso-Coello P, Schünemann HJ, Moberg J, et al. GRADE Evidence to Decision (EtD) frameworks: A systematic and transparent approach to making well informed health care choices. 1: Introduction. *BMJ*. 2016;353:i2016. doi:10.1136/bmj.i2016

16. Eccles MP, Mittman BS. Welcome to implementation science. *Implement Sci*. 2006;1(1):1-3. doi:10.1186/1748-5908-1-1

17. Bauer MS, Damschroder L, Hagedorn H, Smith J, Kilbourne AM. An introduction to implementation science for the non-specialist. *BMC Psychol*. 2015;3(1):1-12. doi:10.1186/S40359-015-0089-9

18. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford university press; 2015.

19. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: The RE-AIM framework. *Am J Public Health*. 1999;89(9):1322-1327. doi:10.2105/AJPH.89.9.1322

20. Hernán MA, Taubman SL. Does obesity shorten life? The importance of well-defined interventions to answer causal questions. *Int J Obes*. 2008;32:S8-S14. doi:10.1038/ijo.2008.82

21. Daniel RM, De Stavola BL, Vansteelandt S. Commentary: The formal approach to quantitative causal inference in epidemiology: Misguided or misrepresented? *Int J*

Epidemiol. 2016;45(6):1817-1829. doi:10.1093/ije/dyw227

22. Cole SR, Stuart EA. Generalizing evidence from randomized clinical trials to target populations: The ACTG 320 trial. *Am J Epidemiol*. 2010;172(1):107-115. doi:10.1093/aje/kwq084

23. Hernán MA, Robins JM. *Causal Inference : What If*. Boca Raton: Chapman & Hall/CRC; 2020.

24. Fleming TR. Surrogate endpoints and FDA's accelerated approval process. *Health Aff (Millwood)*. 2005;24(1):67-78. doi:10.1377/hlthaff.24.1.67

25. Methodology Committee of the Patient-Centered Outcomes Research Institute. Methodological standards and patient-centeredness in comparative effectiveness research: the PCORI perspective. *JAMA*. 2012;307(15):1636-1640. doi:10.1001/jama.2012.466

26. Creswell JW, Creswell JD. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Sage publications; 2017.

27. Khorsan R, Crawford C. External Validity and Model Validity: A Conceptual Approach for Systematic Review Methodology. *Evidence-Based Complement Altern Med*. 2014;2014:1-12. doi:10.1155/2014/694804

28. Alonso-Coello P, Oxman AD, Moberg J, et al. GRADE Evidence to Decision (EtD) frameworks: A systematic and transparent approach to making well informed health care choices. 2: Clinical practice guidelines. *BMJ*. 2016;353:i2089. doi:10.1136/bmj.i2089

29. Parmelli E, Amato L, Oxman AD, et al. GRADE Evidence to Decision (EtD) framework for coverage decisions. *Int J Technol Assess Health Care*. 2017;33(2):176-182. doi:10.1017/S0266462317000447

30. Schünemann HJ, Wiercioch W, Brozek J, et al. GRADE Evidence to Decision (EtD) frameworks for adoption, adaptation, and de novo development of trustworthy recommendations: GRADE-ADOLOPMENT. *J Clin Epidemiol*. 2017;81:101-110. doi:10.1016/j.jclinepi.2016.09.009

31. Moberg J, Oxman AD, Rosenbaum S, et al. The GRADE Evidence to Decision (EtD)

framework for health system and public health decisions. *Heal Res Policy Syst*. 2018;16(1):1-15. doi:10.1186/s12961-018-0320-2

32. Piggott T, Brozek J, Nowak A, et al. Using GRADE evidence to decision frameworks to choose from multiple interventions. *J Clin Epidemiol*. 2021;130:117-124. doi:10.1016/j.jclinepi.2020.10.016

33. Khalifa M, Magrabi F, Gallego B. Developing a framework for evidence-based grading and assessment of predictive tools for clinical decision support. *BMC Med Inform Decis Mak*. 2019;19(1):207. doi:10.1186/s12911-019-0940-7

34. Milat A, Lee K, Conte K, et al. Intervention Scalability Assessment Tool: A decision support tool for health policy makers and implementers. *Heal Res Policy Syst*. 2020;18(1):1-17. doi:10.1186/s12961-019-0494-2

35. Thorpe KE, Zwarenstein M, Oxman AD, et al. A pragmatic-explanatory continuum indicator summary (PRECIS): A tool to help trial designers. *C Can Med Assoc J*. 2009;180(10):E47-57.. doi:10.1503/cmaj.090523

36. Koppenaal T, Linmans J, Knottnerus JA, Spigt M. Pragmatic vs. explanatory: An adaptation of the PRECIS tool helps to judge the applicability of systematic reviews for daily practice. *J Clin Epidemiol*. 2011;64(10):1095-1101. doi:10.1016/j.jclinepi.2010.11.020

37. Moons KGM, Wolff RF, Riley RD, et al. PROBAST: A tool to assess risk of bias and applicability of prediction model studies: Explanation and elaboration. *Ann Intern Med*. 2019;170(1):W1-W33. doi:10.7326/M18-1377

38. Glasgow RE, Klesges LM, Dzewaltowski DA, Estabrooks PA, Vogt TM. Evaluating the impact of health promotion programs: Using the RE-AIM framework to form summary measures for decision-making involving complex issues. *Health Educ Res*. 2006;21(5):688-694. doi:10.1093/her/cyl081

39. Lavis JN, Permanand G, Oxman AD, Lewin S, Fretheim A. SUPPORT Tools for evidence-informed health Policymaking (STP) 13: Preparing and using policy briefs to support evidence-informed policymaking. *Heal Res policy Syst*. 2009;7 Suppl 1(Suppl 1):S13. doi:10.1186/1478-4505-7-S1-S13

40. Stratil JM, Baltussen R, Scheel I, Nacken A, Rehfuess EA. Development of the WHO-INTEGRATE evidence-to-decision framework: An overview of systematic reviews of decision criteria for health decision-making. *Cost Eff Resour Alloc*. 2020;18(1):1-15. doi:10.1186/s12962-020-0203-6

41. Almeida ND, Mines L, Nicolau I, et al. A framework for aiding the translation of scientific evidence into policy: The experience of a hospital-based technology assessment unit. *Int J Technol Assess Health Care*. 2019;35(3):204-211. doi:10.1017/S0266462319000254

42. Atkins D, Chang SM, Gartlehner G, et al. Assessing applicability when comparing medical interventions: AHRQ and the Effective Health Care Program. *J Clin Epidemiol*. 2011;64(11):1198-1207. doi:10.1016/j.jclinepi.2010.11.021

43. Berger ML, Martin BC, Husereau D, et al. A questionnaire to assess the relevance and credibility of observational studies to inform health care decision-making: An ISPOR-AMCP-NPC good practice task force report. *Value Heal*. 2014;17(2):143-156. doi:10.1016/j.jval.2013.12.011

44. Bonell C, Oakley A, Hargreaves J, Strange V, Rees R. Assessment of generalisability in trials of health interventions: suggested framework and systematic review. *BMJ*. 2006;333(7563):346 LP - 349. doi:10.1136/bmj.333.7563.346

45. Bornhöft G, Maxion-Bergemann S, Wolf U, et al. Checklist for the qualitative evaluation of clinical studies with particular focus on external validity and model validity. *BMC Med Res Methodol*. 2006;6:1-13. doi:10.1186/1471-2288-6-56

46. Burford B, Lewin S, Welch V, Rehfuess E, Waters E. Assessing the applicability of findings in systematic reviews of complex interventions can enhance the utility of reviews for decision-making. *J Clin Epidemiol*. 2013;66(11):1251-1261. doi:10.1016/j.jclinepi.2013.06.017

47. Gruen RL, Morris PS, McDonald EL, Bailie RS. Making systematic reviews more useful for policy-makers. *Bull World Health Organ*. 2005;83(6):480. doi:10.1590/S0042-96862005000600020

48. Linan Z, Qiusha Y, Chuan Z, et al. An instrument for evaluating the clinical applicability

of guidelines. *J Evid Based Med*. 2020;14(September):75-81. doi:10.1111/jebm.12416

49. Polus S, Pfadenhauer L, Brereton L, et al. A consultation guide for assessing the applicability of health technologies: A case study. *Int J Technol Assess Health Care*. 2017;33(5):577-585. doi:10.1017/S0266462317000745

50. Cluzeau F, Burgers J, Brouwers M, et al. Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: The AGREE project. *Qual Saf Heal Care*. 2003;12(1):18-23. doi:10.1136/qhc.12.1.18

51. Brouwers MC, Kho ME, Browman GP, et al. AGREE II: Advancing guideline development, reporting and evaluation in health care. *Cmaj*. 2010;182(18):839-842. doi:10.1503/cmaj.090449

52. Brouwers MC, Spithoff K, Kerkvliet K, et al. Development and validation of a tool to assess the quality of clinical practice guideline recommendations. *JAMA Netw open*. 2020;3(5):e205535. doi:10.1001/jamanetworkopen.2020.5535

53. Cambon L, Minary L, Ridde V, Alla F. A tool to analyze the transferability of health promotion interventions. *BMC Public Health*. 2013;13(1):1184. doi:10.1186/1471-2458-13-1184

54. Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: A tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*. 2016;355:1-7. doi:10.1136/bmj.i4919

55. Brouwers MC, Kho ME, Browman GP, et al. AGREE II: Advancing guideline development, reporting, and evaluation in health care. *Prev Med (Baltim)*. 2010;51(5):421-424. doi:10.1016/j.ypmed.2010.08.005

56. Hernán MA. The C-word: Scientific euphemisms do not improve causal inference from observational data. *Am J Public Health*. 2018;108(5):616-619. doi:10.2105/AJPH.2018.304337

57. Vandenbroucke JP, Broadbent A, Pearce N. Causality and causal inference in epidemiology: The need for a pluralistic approach. *Int J Epidemiol*. 2016;45(6):1776-1786. doi:10.1093/ije/dyv341

58. Krieger N, Smith GD. The tale wagged by the DAG: Broadening the scope of causal inference and explanation for epidemiology. *Int J Epidemiol*. 2016;45(6):1787-1808. doi:10.1093/ije/dyw114

59. VanderWeele TJ, Hernán MA, Tchetgen Tchetgen EJ, Robins JM. Re: Causality and causal inference in epidemiology: The need for a pluralistic approach. *Int J Epidemiol*. 2016;45(6):2199-2200. doi:10.1093/ije/dyw162

60. Cole SR, Frangakis CE. The consistency statement in causal inference: A definition or an assumption? *Epidemiology*. 2009;20(1):3-5. doi:10.1097/EDE.0b013e31818ef366

61. Glymour C, Glymour MR. Commentary: race and sex are causes. *Epidemiology*. 2014;25(4):488-490. doi:10.1097/EDE.0000000000000122

62. Fleming TR, Powers JH. Biomarkers and surrogate endpoints in clinical trials. *Stat Med*. 2012;31(25):2973-2984. doi:10.1002/sim.5403

63. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 8. Rating the quality of evidence - Indirectness. *J Clin Epidemiol*. 2011;64(12):1303-1310. doi:10.1016/j.jclinepi.2011.04.014

64. Wasserstein RL, Lazar NA. The ASA's statement on p-values: context, process, and purpose. *Am Stat*. 2016;70(2):129-133. doi:10.1080/00031305.2016.1154108

65. Hunink MGM, Weinstein MC, Wittenberg E, et al. *Decision Making in Health and Medicine: Integrating Evidence and Values*. 2nd ed. Cambridge: Cambridge University Press; 2014. doi:DOI: 10.1017/CBO9781139506779

66. Singh D. Merck withdraws arthritis drug worldwide. *BMJ*. 2004;329(7470):816. doi:10.1136/bmj.329.7470.816-a

67. Poole C. On the origin of risk relativism. *Epidemiology*. 2010;21(1):3-9. doi:10.1097/EDE.0b013e3181c30eba

68. Hernán MA, Vanderweele TJ. Compound treatments and transportability of causal inference. *Epidemiology*. 2011;22(3):368-377. doi:10.1097/EDE.0b013e3182109296

69. Bareinboim E, Pearl J. Causal inference and the data-fusion problem. *Proc Natl Acad Sci*.

2016;113(27):7345-7352. doi:10.1073/pnas.1510507113

70. Cambon L, Minary L, Ridde V, Alla F. A tool to facilitate transferability of health promotion interventions: ASTAIRE. *Sante Publique (Paris)*. 2014;26(6):783-786. doi:10.3917/spub.146.0783

71. Viswanathan M, Ansari MT, Berkman ND, et al. Assessing the Risk of Bias of Individual Studies in Systematic Reviews of Health Care Interventions. In: *Methods Guide for Effectiveness and Comparative Effectiveness Reviews [Internet]*. Rockville (MD); 2012.

72. Baker A, Potter J, Young K, Madan I. The applicability of grading systems for guidelines. *J Eval Clin Pract*. 2011;17(4):758-762. doi:10.1111/j.1365-2753.2011.01693.x

73. Siering U, Eikermann M, Hausner E, Hoffmann-Eßer W, Neugebauer EA. Appraisal tools for clinical practice guidelines: A systematic review. *PLoS One*. 2013;8(12). doi:10.1371/journal.pone.0082915

74. Graham ID, Calder LA, Hébert PC, Carter AO, Tetroe JM. A comparison of clinical practice guideline appraisal instruments. *Int J Technol Assess Health Care*. 2000;16(4):1024-1038. doi:10.1017/S0266462300103095

75. Wells GA, Shea B, O'Connell D al, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. 2000.

76. Lewis D. *Counterfactuals*. John Wiley & Sons; 2013.

77. Höfler M. Causal inference based on counterfactuals. *BMC Med Res Methodol*. 2005;5:1-12. doi:10.1186/1471-2288-5-28

78. Brodersen J, Schwartz LM, Heneghan C, O\textquoterightSullivan JW, Aronson JK, Woloshin S. Overdiagnosis: what it is and what it isn't. *BMJ Evidence-Based Med*. 2018;23(1):1-3. doi:10.1136/ebmed-2017-110886

79. Ioannidis JPA. The Proposal to Lower P Value Thresholds to .005. *JAMA*. 2018;319(14):1429-1430. doi:10.1001/jama.2018.1536

80. Noordzij M, Van Diepen M, Caskey FC, Jager KJ. Relative risk versus absolute risk: One cannot be interpreted without the other. *Nephrol Dial Transplant*. 2017;32(February):ii13-

ii18. doi:10.1093/ndt/gfw465

6.11. TABLES

Table 6.1. Frameworks for appraising applicability of studies

Framework and/or First Author		Journal	Framework Focus	Year of Publication
AGREE I AGREE II	Cluzeau et al.[50] Brouwers et al.[51]	*Quality and Safety in Health Care* *Canadian Medical Association J*	Quality of CPG	2003 2010
AGREE-REX	Brouwers et al.[52]	*JAMA Network Open*	Quality of CPG recommendations	2020
ASTAIRE	Cambon et al.[53]	*BMC Public Health*	Transferability of health promotion interventions	2013
EVAT	Khorsan et al.[27]	*Evidence-Based Complementary and Alternative Medicine*	Evidence for clinical decision-making	2014
GRADE	Guyatt et al.[11]	*British Medical Journal*	Quality of evidence and strength of recommendations	2008
GRADE EtD Clinical recommendations Coverage decisions Diagnostic/screening tests Health system and public health Multi-intervention comparisons	Alonso-Coello et al.[15,28] Parmelli et al.[29] Schünemann et al.[30] Moberg et al.[31] Piggott et al.[32]	 *British Medical Journal* *Int J of Tech Assessm in Health Care* *Journal of Clinical Epidemiology* *Health Research Policy and Systems* *Journal of Clinical Epidemiology*	Evidence usage in a structured and transparent way to inform and adapt clinical and public health decisions	2016 2016 2017 2017 2018 2021
GRASP	Khalifa et al.[33]	*BMC Medical Informatics and Decision Making*	Predictive tools for clinical decision support	2019
ISAT	Milat et al.[34]	*Health Research Policy and Systems*	Decision support tool for health policy makers and implementers	2020
PRECIS	Thorpe et al.[35]	*Canadian Medical Association Journal*	Pragmatic vs. exploratory trials for trial designers	2009
PR-Tool	Koppenaal et al.[36]	*Journal of Clinical Epidemiology*	Applicability of individual and SR of trials	2011

PROBAST	Moons et al.[37]	*Annals of Internal Medicine*	Risk of bias and applicability of prediction model studies	2019
RE-AIM	Glasgow et al.[19,38]	*American Journal of Public Health Health Education Research*	Evaluate and report on internal and external validity, and impact of health promotion programs	1999 2006
RoB 2	Sterne et al.[7]	*British Medical Journal*	Risk of bias in randomized trials	2019
RoBINS-I	Sterne et al.[6]	*British Medical Journal*	Risk of bias in non-randomised studies of interventions	2016
STP	Lavis et al.[39]	*Health Research Policy and Systems*	Applicability of the findings of a systematic review	2009
WHO-INTEGRATE EtD	Stratil et al.[40]	*Cost Effectiveness and Resource Allocation*	Decision criteria for health decision-making	2020
Almeida et al.[41]		*Int J of Tech Assessm in Health Care*	Translation of HTA evidence into policy	2019
Atkins et al.[42]		*Journal of Clinical Epidemiology*	Applicability when comparing medical interventions for SR	2011
Berger et al.[43]		*Value in Health*	Relevance and credibility of observational studies for health care decision-making	2014
Bonell et al.[44]		*British Medical Journal*	Generalizability in trials of health interventions	2006
Bornhoft et al.[45]		*BMC Medical Research Methodology*	Evaluation of clinical studies on external and model validity	2006
Burford et al.[46]		*Journal of Clinical Epidemiology*	Applicability of findings in systematic reviews of complex interventions for SR	2013
Green et al.[14]		*Evaluation and the Health Professions*	Relevance, generalizability, and applicability of research	2006
Gruen et al.[47]		*Bulletin of the World Health Org*	Generalizability of studies in LMIC for SR	2005
Linan et al.[48]		*Journal of Evidence-Based Medicine*	Clinical applicability of CPG	2020

Polus et al.[49]	*Int J of Tech Assessm in Health Care*	Applicability of a technology in the context of HTA	2017

Notes. CPG = clinical practice guidelines HTA = health technology assessment; LMIC = low- and middle-income countries; SR = systematic review;

Table 6.2. VICORT criteria definition and relation to other epidemiological concepts

VICORT Criteria	Definition	Related criteria from literature synthesis	Relation to other epidemiological frameworks or concepts
Validity	Methods are appropriate for internal validity of results: • Experimental evidence generated is not subject to randomization, blinding, protocol deviation, missing data, or measurement issues. • Observational evidence generated is not subject to confounding, information, and endogenous selection biases.	Confounding Consistency of results Dose-response gradient Precision Reporting bias Risk of bias Strength and level of evidence	Quality assessment tools, e.g., - Cochrane risk-of-bias tool[7] - Risk of bias in non-randomised Studies of Interventions)[6] - Newcastle-Ottawa Scale[75]
Indication-informativeness	Study methods provide clinicians with evidence to determine a clinical indication in specific individuals. Informativeness *for a clinical indication* requires a well-defined intervention whose effect can be identified from the study results, i.e.: 1. A trial of an intervention (experimental study) in specific/eligible individuals; **OR** 2. An observational study of an exposure where: A) A well-defined intervention for specific individuals (those with the exposure) exists, **AND** B) That the effect of this well-defined intervention be correctly identified (independent effect of the intervention); **OR** 3. An observational study where there is an intervention on specific individuals and where *absolute* results for outcomes are explicitly reported. (*informativeness for the outcome of an intervention* criterion – allows contrast between intervention in specific individuals and envisioned natural history under no intervention in those individuals.)	*None*	Counterfactuals[76,77] Well-defined intervention, consistency assumption of causal inference[23,60]
Clinical relevance	Primary outcome of the study is clinically relevant, i.e., the outcome is at a minimum clinically informative, and ideally, patient centered.	Directness Relevance of outcomes	Surrogate outcomes[62] Overdiagnosis[78] Patient-centered outcomes research[25]
Originality	**Significance**. Study results achieve clinical (not only statistical) significance (e.g., a relevant magnitude of effect); **AND** **Novelty**. Study results are novel when compared to current evidence base and practice.	Comparison intervention Intervention characteristics Magnitude (effect size) and trade-offs of harms and benefits	Clinical *vs.* sole statistical significance Dichotomization *vs.* magnitude of effect and confidence intervals[64,79]

Risk-benefit comprehensiveness	Overall benefits of changing an indication (either the intervention or the population of individuals in which the intervention is indicated) comprehensively outweigh the risks.	Magnitude and trade-offs of harms and benefits	Net benefit - Generic health state measures Relative vs. absolute measures[67,80]
Transposability	The clinical indication/intervention is implementable and (cost-) effective in the specific practice setting.	Acceptability and feasibility Context and resources for application Cost and cost-effectiveness Equity and ethics Monitoring/audit and support tools Representativeness of patients and populations Scope of practice and actions Sustainability Values, beliefs, preferences priority	Generalizability and transportability[22,68,69] Cost-effectiveness analysis[18]

6.12. FIGURES

Figure 6.1. Existing frameworks for the appraisal of applicability according to evidence type and target user

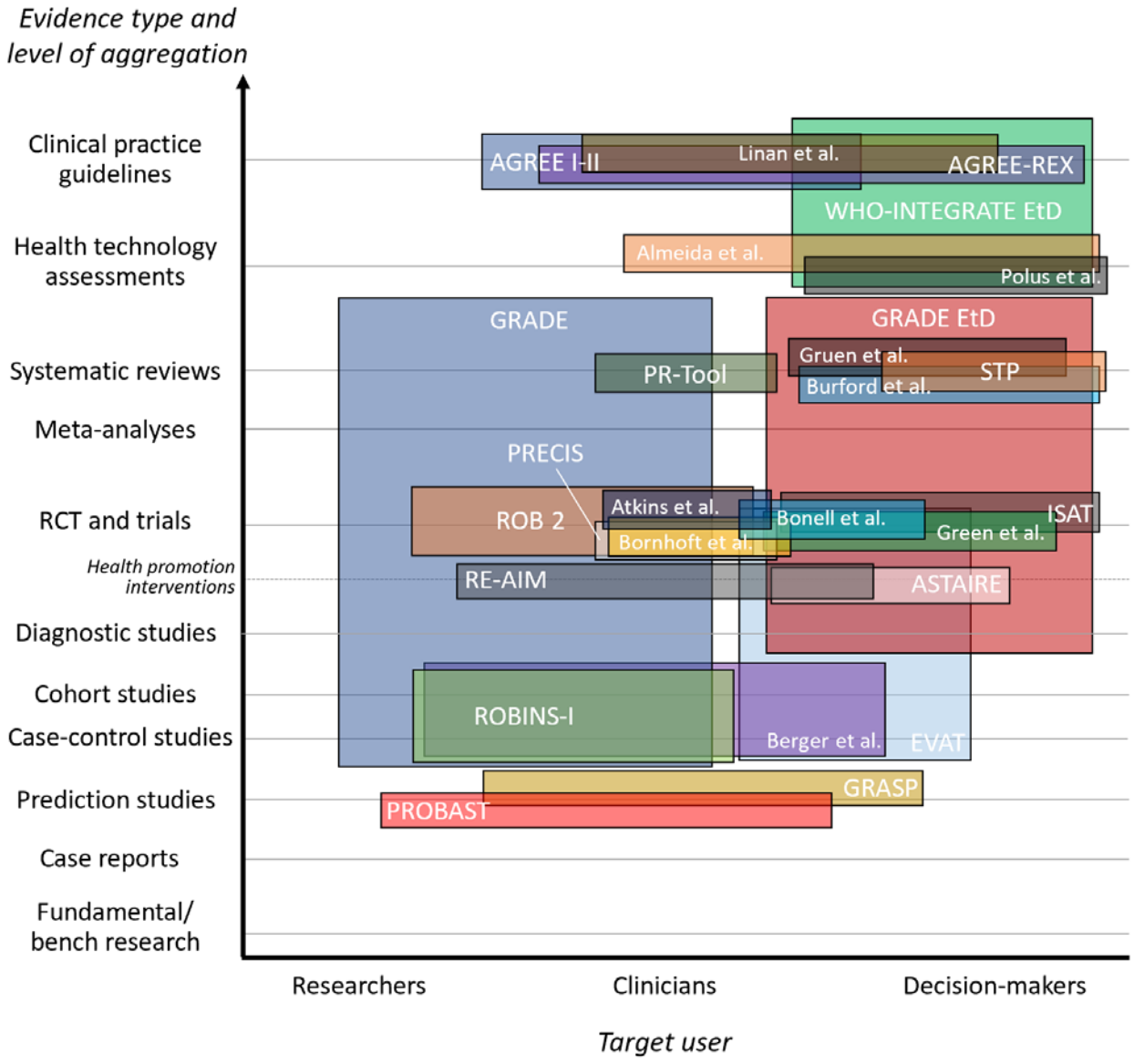
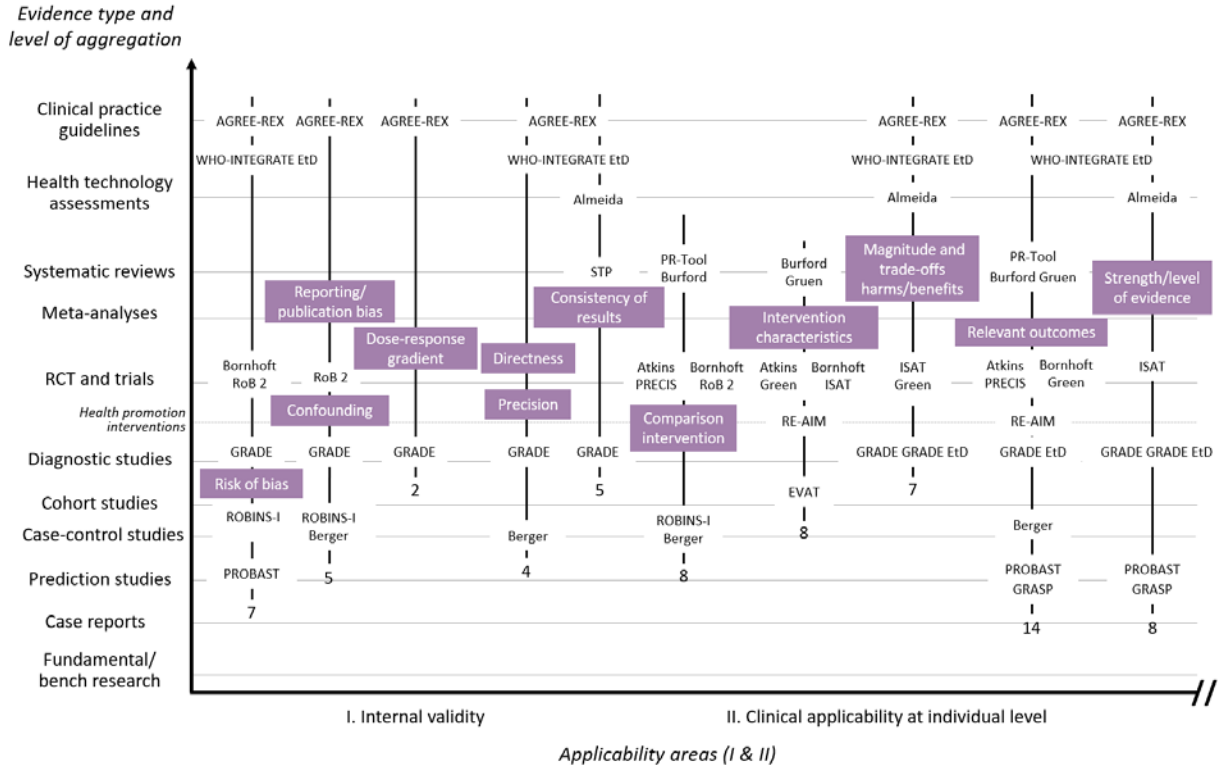


Figure 6.2. Criteria used to appraise applicability by framework, frequency, and according to evidence type and applicability domain (part 1)



Note. The number under each vertical line indicates the count of frameworks (n=26) featuring this criterion.

Figure 6.2. Criteria used to appraise applicability by framework and according to evidence type and applicability domain (part 2)

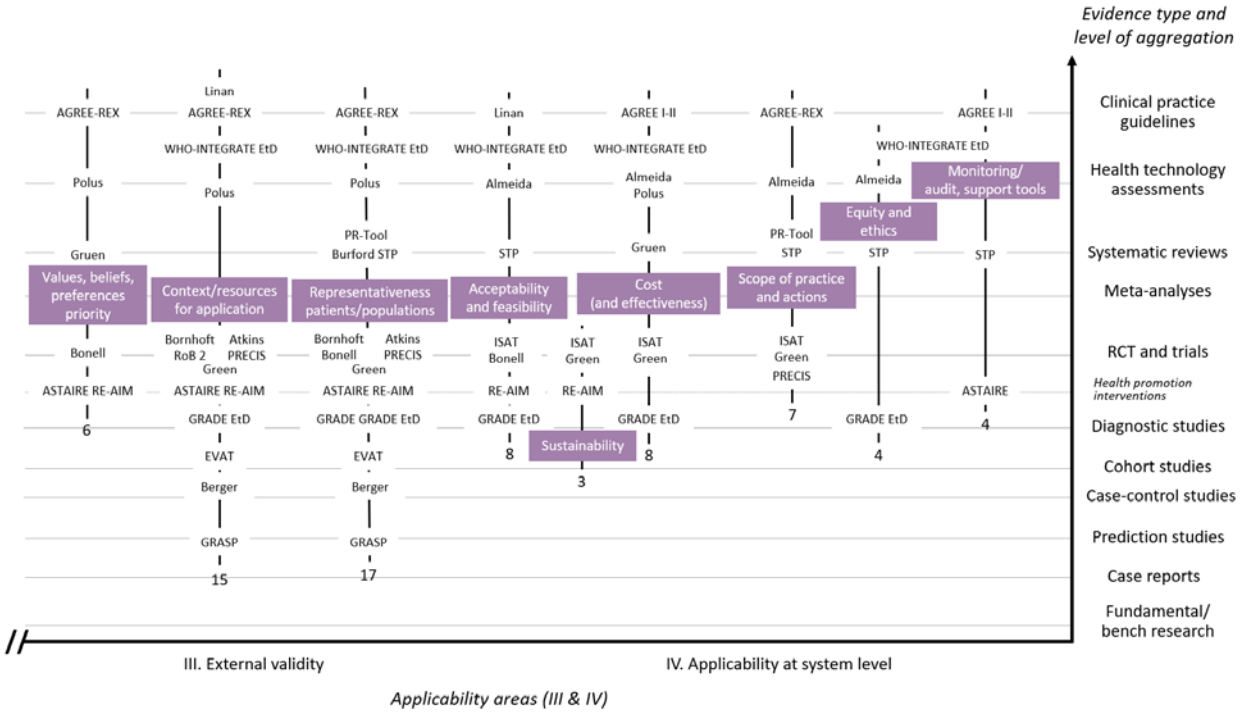


Figure 6.3. Framework for appraising clinical applicability of studies (FrACAS)

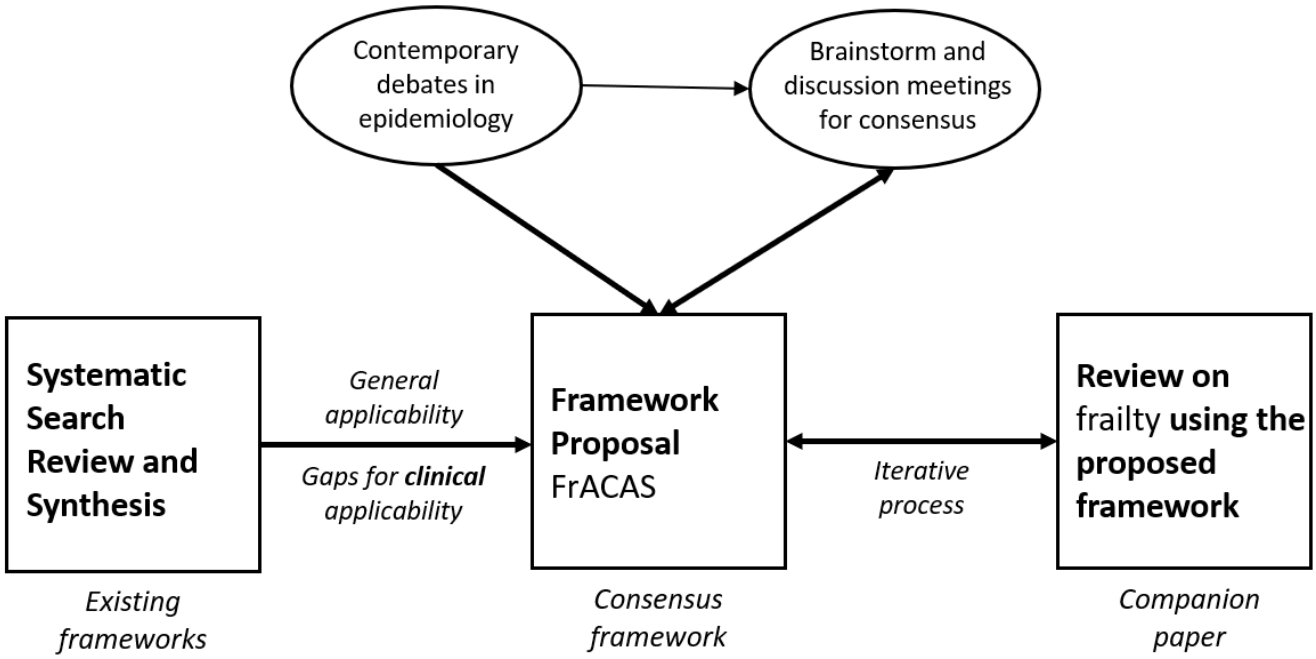
6.13. SUPPLEMENTAL MATERIAL

Supplementary Methods.

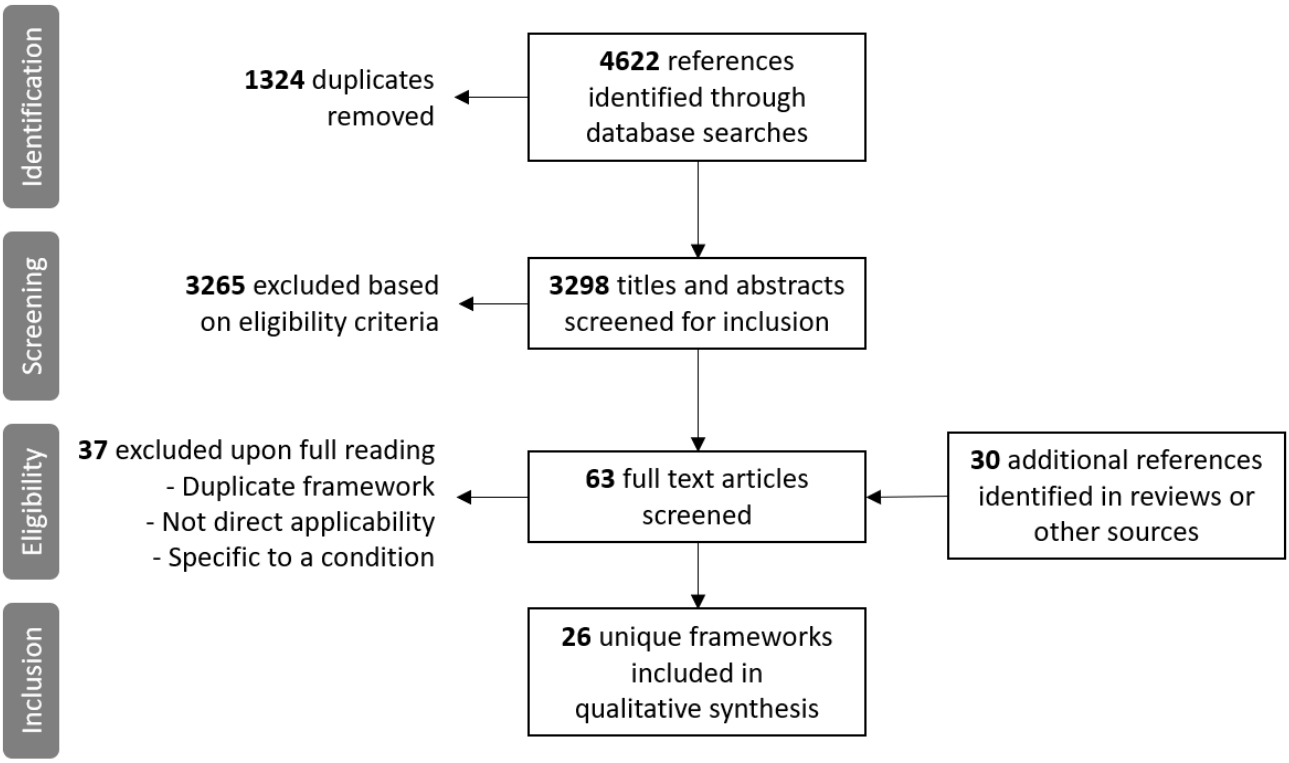
Search query for Embase using OVID
(("clinical practice" or decision or applica* or "clinical impact" or transferability or translat* or "external validity" or generalisability) and (apprais* or asses* or grad* or analy* or "systematic reviews" or evaluat* or factor*) or (framework or recommendation or consensus or guide or checklist or tool* or instrument or program or questions)).ti. and english.lg.

Search query for PubMed
("clinical practice" or "evidence" or "decision" or "applica*" or "clinical impact" or “transferability” or “translation” or “implementation” or “external validity”) and ("apprais*" or "assess*" or "grad*" or “analyse” or “analyze” or “systematic reviews”) ("framework" or "recommendation" or "consensus" or "guide" or "checklist" or "tool*" or "instrument*" or “program”) and english[Language]

Supplementary Figure 6.1. Process and inputs for the development of the Framework for Appraising the Clinical Applicability of Studies (FrACAS)



Supplementary Figure 6.2. Flowchart for selection of articles



Chapter 7. Manuscript 4: The State of frailty in research: a mapping review of its clinical applicability to practice

7.1 PREFACE

Manuscript 1 examined whether there was a basis for categorizing older adults. Manuscript 2 examined whether a specific categorization using frailty indices was sufficiently reliable. Manuscript 3 proposed an appraisal framework to determine the clinical applicability of studies. In this final manuscript, I built upon the previous chapters to investigate the clinical applicability of frailty in the recent literature. I conducted a scoping review that touches upon the three facets previously examined or discussed which are central to clinical applicability:

1. CATEGORIZATION AND IDENTIFICATION. What does frailty, as used in the current literature, measure or concern itself with, and how similar are the various definitions and measures of frailty?
2. FUNCTIONAL CLASS. How is frailty used, under what epidemiological or clinical function?
3. ACTION RELATION. Can the measurement of frailty lead to a clinical action?

Examining these facets together may assist in uncovering whether and why frailty—and other constructs—can fulfill the clinical imperative of materially improving health outcomes. It may also allow researchers and clinicians to improve upon existing definitions and uses of frailty to improve translation into practice.

***

This manuscript has been submitted to Ageing Research Reviews.

7.2 TITLE PAGE

The State of frailty in research: a mapping review of its clinical applicability to practice

Short title

State of frailty in research and practice – Nguyen et al.

Authors and affiliations

Quoc Dinh Nguyen, MD, MA, MPH,[a,b,c] Erica M. Moodie, PhD,[c] Philippe Desmarais, MD, MHSc,[a,b] Marie-France Forget, MD, MSc,[a] Han Ting Wang MD, MSc, [d] Mark R. Keezer, MDCM, PhD,[b,e] Christina Wolfson, PhD[c,f,g]

a. Division of Geriatrics, Department of Medicine, Centre hospitalier de l'Université de Montréal, Canada;
b. Centre de recherche du Centre hospitalier de l'Université de Montréal, Canada
c. Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, Canada;
d. Department of Internal Medicine and Critical Care Medicine, Hôpital Maisonneuve-Rosemont, Montreal, Canada
e. Departments of Neurosciences & Social and Preventative Medicine, Université de Montréal, Canada
f. Department of Medicine, McGill University, Canada;
g. Neuroepidemiology Research Unit, Research Institute of the McGill University Health Centre, Canada.

Corresponding author

Quoc Dinh Nguyen
1000, rue Saint-Denis, Montreal, Quebec, Canada H2X 0C1
quoc.nguyen@mail.mcgill.ca

Manuscript word count

Abstract word count: 272
Word count: 3733 words
Tables: 3
Figures: 2
References: 56

7.3. ABSTRACT

Research on frailty has expanded in the last decade, but direct evidence supporting its implementation in clinical practice may be limited. This mapping review synthesizes the contexts-of-use and overall clinical applicability of recent frailty research. We sampled 476 articles from those published on frailty in PubMed and EMBASE in 2017-2018, of which 150 articles were fully appraised fully for the contexts-of-use, definitions, and interventions. A clinical applicability framework was used to classify articles as practice-changing, practice-informing, or not practice-informing. Of 476 articles, 150 (31%) used frailty in functions that could inform a clinical indication: predictor or mediator (n = 125, 26%), selection criterion (15, 3%), and effect modifier (10, 2%). Articles spanned all health disciplines, and cohort studies comprised 137 (91%) studies and trials 13 (9%). Thirty-eight frailty definitions using varied cut-offs and a wide range of interventions were identified. Among all articles, 63 (13%) articles were practice-informing, 11 (2%) potentially practice-changing, and 1 (0.2%) clearly practice-changing. Lack of well-defined intervention and identifiable effect (96%) or originality (83%) were predominant reasons reducing applicability. Only a minority of recent frailty research provides direct evidence of applicability to practice. Future research on frailty should focus on translating frailty, as a risk factor, into a clinical indication and address definition ambiguity.

Keywords

Frailty, clinical practice, risk factor, interventions, epidemiology

7.4. INTRODUCTION

The number of older adults with frailty will increase in the next decades.[1] Frailty, a state of increased vulnerability and decreased resilience to stressors, has been associated with morbidity, functional decline, and mortality.[2] In keeping with population aging, research on frailty has robustly expanded over the past 15 years with more than 600 publications in 2017 alone[3]. Two major frameworks underpin the current understanding of frailty:[4] the Fried physical frailty phenotype (PFP), and the Rockwood deficit-accumulation frailty index (FI).[5–7] Even if defined using various instruments, frailty has been consistently associated with adverse outcomes,[8–10] leading to calls from the clinical community to include frailty assessment in the management of older adults.[11–14]

However, concerns regarding the paucity of evidence for the direct clinical utility and applicability of frailty have been raised by many.[15–17] Despite the growing body of literature, the evidence base supporting the use frailty to inform current clinical care decisions may be limited. Only a minority of studies have examined frailty as a means to improve or alter clinical decision making.[9,10] Moreover, most studies used frailty for risk stratification for adverse outcomes, and association studies vastly outnumbered intervention studies.[3,10] Previous mapping and systematic reviews have described the prevalence of frailty,[18] categorized instruments used to measured frailty as exposure[9,19] or outcome[20] and characterized the domains assessed[21,22] and agreement[23–25] between frailty operationalizations. None have specifically investigated the applicability of frailty to clinical practice. To clinicians overwhelmed by the rapidly growing literature, a crucial question may be to understand how and/or whether frailty may directly alter current clinical practice.

In this study, we conduct a mapping review to better characterize the overall clinical applicability of recent frailty research. We first describe and then map the contexts-of-use of frailty

in work published in 2017-2018, with an emphasis on its definitions and operationalizations. Second, we use a framework, which we developed drawing from the ideas of well-defined intervention[26,27] and patient-centered outcomes research,[28] to classify and quantify how much of the recent frailty research output informs or changes clinical practice. Finally, we analyze researchers' perception and statement of clinical usefulness of frailty and synthesize reported frailty-related interventions.

7.5. METHODS

7.5.1. Data sources and searches

The literature on frailty has vastly expanded in the last decade making appraisal of all articles infeasible. We focused our mapping review[29] on the literature published in 2017 or 2018 and used systematic sampling to select articles as described below. Figure 7.1A shows the flowchart for the article selection process.

First, we identified all articles about clinical frailty. We searched MEDLINE (PubMed) and EMBASE (Ovid) databases for articles published on frailty in English or French, from January 1, 2017 to December 31, 2018 (the full search strategy is detailed in Supplementary Methods 1).

7.5.2. Study selection

After removing duplicates, one reviewer (QDN) examined titles and abstracts to remove articles indexed in 2017 or 2018 but published outside that timeframe, replies to the editor, and papers that were not scientific journal articles. Articles were then reviewed to exclude articles that did not significantly feature *clinical* frailty (as reported at the second step of Figure 7.1A) or did not use a previously published measure of frailty. This latter step was done by one reviewer (QDN), and a random sample of 5% of articles were validated by a second reviewer (MFF).

From all remaining articles about clinical frailty, our goal was to assess the primary evidence for clinical applicability of frailty, i.e., to inform a clinical indication for individuals with frailty. To achieve this, articles were randomly ordered, and pairs of independent reviewers (PD, MFF, QDN, HTW) followed this order to classify articles by the primary function of frailty (predictor or mediator, selection criterion, effect modifier, outcome, study population descriptor, confounder—described in Supplementary Methods 1) or by article type (measurement of frailty,

review/guideline/consensus, editorial/commentary, other). This order was followed until 150 articles were identified that were reports from observational studies or clinical trials using frailty in one of the following four functions: predictor or mediator, effect modifier, or primary selection criterion. Since they could not directly inform a clinical indication for frailty, we enumerated, but did not fully appraise, other frailty functions (outcome measure, confounder) and article types in order to quantify the total number of articles classified.

7.5.3. Data extraction

We extracted the following study characteristics for the 150 fully appraised articles: journal and year of publication, study design, sample size, age, percentage of women, setting, study population, intervention studied (when the study was a trial), and primary outcome. We also extracted data on the operationalization of frailty: distribution and prevalence of frailty, definition and cut-offs used, domains (activity-exhaustion-energy, cognition, function and disability, health care use, laboratory values, medical conditions-symptoms, mood-psychiatric, nutrition-weight, physical function, physiological function, subjective health, and vital signs), and number of items considered in definitions of frailty.

7.5.4. Data synthesis and analysis of clinical applicability

We developed an appraisal framework, summarized in Supplementary Methods 2, which draws from ideas of well-defined intervention[26,27] and patient-centered outcomes research.[28] Using this framework, we appraised the potential for clinical usefulness and applicability of articles on frailty by classifying articles as being practice changing (and informing), practice informing (but not practice changing), or not practice informing. Figure 7.1B summarizes the appraisal process. Classification was determined by four intermediate criteria: (A) indication-informativeness, (B) clinical relevance, (C) originality, and (D) risk-benefit comprehensiveness. To be classified as

practice changing, articles had to satisfy all four criteria; to be classified as practice informing, articles had to fulfill criteria A (indication-informativeness) and B (clinical relevance); remaining articles were classified as not practice informing. Our appraisal of each intermediate criterion was "yes", "no", "uncertain"; when an intermediate criterion was deemed uncertain, we classified the article as "potentially" practice informing or practice changing. Supplementary Methods 2 provides detailed definitions, descriptions, and rationale for the intermediate criteria and classification. Finally, we determined whether articles stated a clinical usage of frailty by reading the article in full, and we extracted interventions described in relation to frailty. Data extraction and appraisal were carried out by pairs of independent reviewers with methodological and aging-related expertise (PD, MFF, QDN, HTW). Discrepancies were resolved by consensus.

7.5.5. Statistical analysis

Descriptive statistics are presented using means and medians for continuous variables and counts and percentages for categorical variables. Individual articles were considered as the unit of analysis: unweighted summary measures are reported. We calculated unweighted kappa statistics for reliability in study selection and appraisal, with values between 0.21-0.4, 0.41-0.6, 0.61-0.8, 0.81-1 considered respectively fair, moderate, substantial, and almost perfect.[30] For reliability in appraisal of clinical applicability and clinical usage, we additionally calculated the prevalence-adjusted bias-adjusted kappa (PABAK) because of low prevalence in some categories.[31]

7.6. RESULTS

7.6.1. Study selection and appraisal reliability

Our search returned 8841 articles. We identified 5240 unique original journal articles, of which 2807 pertained specifically to clinical frailty (kappa 0.93 [0.88, 0.97]). A total of 476 articles were appraised to include 150 articles that had the potential to inform clinical indications for individuals with frailty (as frailty was used as a predictor or mediator, selection criterion, or effect modifier; kappa 0.78 [0.74, 0.82]). The overall kappa for appraisal of clinical applicability and statement of usage was 0.74 (0.68, 0.81). Agreement, kappa, and PABAK for each appraisal step are reported in Supplementary Table 7.1. Full extraction and appraisal results are available online [*Full Extraction and Appraisal Results.xlsx*].

7.6.2. Frailty functions and contexts-of-use

Figure 7.2A shows the distribution of articles by function and article type. Of the 476 articles, 150 (31%, of total) investigated frailty in functions that had the potential to inform clinical indication: most of these articles used frailty as a predictor or mediator (n = 125, 26%), 15 (3%) used frailty as a selection criterion, and 10 (2%) as an effect modifier. Out of the remaining 326 (69%) articles that could not provide primary evidence to inform clinical indication, 68 (14%) used frailty as an outcome measure, 170 (36%) were not reports of primary data of which 123 (26%) were reviews, guidelines, or consensus articles and 47 (10%) were editorials, comments, or conceptual articles.

Table 7.1 reports study characteristics for the 150 articles with the potential to inform clinical indication. Articles were published in diverse journal disciplines, with aging and geriatrics, medical subspecialties, surgery and anesthesia comprising 68% of articles. Most articles were cohort studies (n = 137, 91%), and 13 (9%) were trials. Median sample size was 476 (IQR: 151-1891; range: 5–962,913), mean age was 72.6 years (SD, 9.8), and women comprised on average

53% (19) of study populations. The setting of participant enrollment was community-dwelling in 53 (35%) studies, institutions in 2 (1%), and mixed/hospital-based in 95 (63%). When study enrollment was hospital-based, surgery (n = 34, 41%) and cardiology (n = 15, 18%) constituted the majority of study disciplines. Primary outcomes were clinical in 106 (71%) studies, health care services-related in 26 (17%), and specifically patient-centered in 18 (12%).

7.6.3. Definitions and operationalizations of frailty

Excluding studies where frailty was the selection criterion, the median prevalence of frailty was 19% (IQR, 12%–39%; range, 1%–84%). Table 7.2 summarizes the definitions and operationalizations used in the 150 articles. A total of 38 different definitions were used: the Fried physical frailty phenotype and modifications were the most commonly used (n = 45, 28%), followed by FI-derived definitions of frailty (n = 35, 22%) and the Clinical frailty scale (CFS, n = 16, 10%).[32] Among studies using FI-derived definitions of frailty, the mean FI was 0.24 (SD, 0.11), the median number of items used was 32 (IQR, 11–41; range, 3–65). Cut-offs for FI-derived definitions varied substantially by implementation: mean cut-off was 0.29 (SD, 0.08; range 0.21-0.41). Similarly, there was variation in cut-offs used for the CFS, grip strength, gait speed, and Edmonton frail scale.[33] Out of 12 frailty domains assessed, the mean number assessed across all definitions of frailty was 3.6 (SD, 2.1) and 4.8 (SD, 2.7) for FI-derived definitions. Supplementary Figure 7.1 shows the proportion of inclusion in frailty definitions for each health domain.

7.6.4. Clinical applicability

Figure 7.2B shows clinical applicability of frailty articles based on the random sample of 476 articles, by function of frailty and study design. Of the sample, 413 (87%) articles were classified as not practice informing: 326 (69%) did not use frailty as a predictor, mediator, effect modifier, or selection criterion, and 87 (18%) were articles where frailty was used a predictor or effect

modifier but did not inform a clinical indication. Sixty-three articles (13%) were classified as practice informing, of which 52 (11%) were classified as not practice changing. Overall, eleven (2%) articles could alter practice: 10 (2%) articles were classified as potentially practice changing, and 1 (0.2%) was classified as changing practice.

Within the 150 fully appraised articles, the two most important reasons for lack of clinical applicability were absence of well-defined intervention and identifiable effect (96% of appraised studies; part of criterion A) and originality (83%). Conversely, 36% of articles were clinically informative by explicitly reporting absolute outcome results for an intervention on individuals with frailty. Detailed results for the appraisal of intermediate criteria are reported in Supplementary Table 7.2. Of the 13 trials, 9 did not report results deemed original, whereas 4 reported results that were considered potentially original. When determined solely by statistical significance, 119 (79%) of 150 articles reported positive results.

7.6.5. Statement of clinical usage and clinical interventions for frailty

Among articles fully appraised, 100 (67%) articles reported statements suggesting or recommending the usage of frailty to inform or alter clinical practice. Supplementary Table 7.3 summarizes interventions for frailty reported in 62 articles. Exercise and multidomain preventive intervention (n = 17) and general tailoring and targeting of interventions (n = 11) were the most frequently described, but there was a very wide spectrum of interventions ranging from the broad comprehensive geriatric assessment (CGA) and management to the narrow proper dosing of anticoagulant therapies.

7.7. DISCUSSION

Our findings provide a means to map and contextualize the rapidly expanding research output on frailty. Frailty is currently investigated across all health research disciplines, with the most prototypical research article published in 2017-2018 reporting results from a hospital-based cohort undergoing surgery where frailty is studied as a predictor of mortality. No dominant definition nor intervention for frailty were identified in 150 randomly sampled articles: 38 different definitions of frailty, most with various cut-offs, were used, along with a wide range of potential interventions. The many coexisting definitions and interventions for frailty may underlie and enable its broad appeal to both the research and clinical communities, as indicated by the abundant number of reviews, editorials, or consensus articles we found across all disciplines. However, when examined through the lens of clinical applicability, only a third of all publications about clinical frailty investigated it in a function informative for clinical applicability; more than a third of publications did not report on primary data. Only 13% of articles were classified as practice informing, 2% as potentially practice changing, and 0.2% as clearly practice changing. Most studies did not feature a well-defined intervention and identifiable effect for frailty or original findings. The low proportion of studies with clinical applicability may explain why calls for the incorporation of frailty in practice have not always been heeded.[13,17,34]

Drawing from our framework and from previous reviews [9,10,16,18–20,23,35], we believe two major themes currently impede the ability of frailty research to alter the practice of clinicians when caring for individuals with frailty: translating frailty as a risk factor to clinical intervenability, and using frailty for prevalence, for assessment of generalizability, and as a selection criterion when the definition of frailty is itself ambiguous.

7.7.1. Frailty as a predictor: translating a risk factor into a clinical indication

The majority of recent studies on frailty are association studies showing relationships with adverse outcomes. However, knowing that individuals with frailty have worse outcomes than those without frailty does not, by itself, entail an intervention (or the possibility of one) to improve outcomes in those with frailty.[36] Prediction of adverse outcomes by frailty can only alter clinical practice, first, if the risk of adverse outcomes for those with frailty exceeds potential benefits in the context of an otherwise indicated intervention (e.g., surgery); or second, in the absence of decision-making for such an intervention, if specific or well-defined interventions exist that can reverse or mitigate frailty and the adverse outcomes due to it (i.e., there is a well-defined intervention and identifiable effect for frailty).

Pertaining to the first issue, our mapping review suggests that most predictive studies where frailty is studied in the context of an intervention compare individuals *with* and *without* frailty, rather than comparing individuals, *all* with frailty, under the scenarios of intervention or no intervention. We did not identify any study reporting the appropriate comparison, that is, either by explicitly reporting the contrast of intervened versus not intervened or by showing that frailty was a qualitative effect measure modifier of an intervention (rather than a simple predictor of adverse outcomes). Nonetheless, some studies were clinically informative as they reported absolute outcome results in individuals with frailty who received an intervention thus allowing the clinical extrapolation and contrast between outcomes under intervention versus no intervention. Though there is generally greater morbidity and mortality in those with frailty following an intervention, there is currently little direct or conclusive evidence that individuals with frailty (or at what level of frailty) would not benefit from an intervention [17,35].

Second, frailty could lead to changes in practice if well-defined frailty-specific interventions exist that can reverse it and mitigate its adverse outcomes. We did not identify such evidence in our mapping review, in part due to the very low proportion of trials (n=13). Two recent reviews specifically synthesizing evidence to prevent or manage frailty suggest that physical activity and prehabilitation interventions may reduce frailty levels, with evidence quality assessed as very low or low in one study,[37] and moderate to good in another.[38] It remains unclear, however, how much of this reduction of frailty would translate into reduction of the actual adverse outcomes associated with frailty in the bulk of prediction studies.

7.7.2. Definition ambiguity: prevalence, generalizability, and selection criterion

The lack of a consensus definition for frailty has been previously reported and discussed,[39,40] but its full clinical implications may not have been fully acknowledged. Clinical entities, such as frailty, can be used, descriptively, to report prevalence and, clinically, to assess generalizability of study results and as a selection criterion for an intervention. The ongoing low agreement of frailty definitions[4,23–25,41] precludes generating applicable evidence about it as a coherent whole in any of these three functions.

First, in line with previous work, we found a wide range of frailty prevalence in studies, which may be attributable to different populations under study,[42] but also to different definitions or operationalizations of frailty used. Additionally, our findings suggest that different *cut-offs* for similar definitions of frailty may compound the variability. Due to variation in definitions and cut-offs, it may be arduous, if not impossible, in practice to keep track of the nature and the level of the "frailty" that is measured when its prevalence is reported.

Second, assessing generalizability of study findings is important for clinical applicability to ensure that evidence applies to older populations, especially in randomized controlled trials due to the relative exclusion of older, multimorbid, and frail participants.[43,44] Observational studies,[45] secondary analyses from trials,[46,47] and practice guidelines have used frailty as a means to appraise generalizability, with some suggesting differential management by frailty status.[48] However, since various definitions and cut-offs coexist, appraising generalizability and applicability of evidence to individual patients using frailty can only be clinically implemented with difficulty.

Third, beyond generalizability, the ultimate objective of a clinical entity is to provide a clinical indication, i.e., to be coupled with an intervention that is beneficial. As such, the identification of the clinical entity of frailty should entail an intervention that will be beneficial to those so identified (the *specific* population with frailty). Yet, myriad definitions of frailty confuse attempts to reliably identify which clinical intervention could best follow from having "frailty." Conversely, myriad purported interventions for frailty confuse attempts to identify who (under which definition of frailty) would *most* benefit from any given well-defined clinical intervention.[49] Since frailty may encompass both single domain and multiple domains,[21] and pertain to community-dwelling adults[18,38,50] and to acute care settings,[10,16] interventions can be both targeted (e.g., exercise training) and broad (e.g., CGA), with long and short timescales. From a clinical standpoint, ambiguity in frailty definitions confounds age-related domains (e.g., chronic conditions, disability, cognition, social support) and opportunities for intervention.

The ambiguity of frailty definitions appears to be overcome by purposely entertaining a wide basket of potential *interventions* for frailty, namely by using frailty as a screening tool[51] for more extensive clinical management, including CGA and ensuing multidimensional interventions.[12] Ambiguity in definition (identifying who will benefit from specific frailty

interventions) can also be circumvented by considering, *as frailty-specific,* interventions that are beneficial to a wide population. This may underlie some recommended interventions for frailty such as exercise training, addressing polypharmacy, and holistic review of chronic conditions, that are already recommended notwithstanding frailty.[12,52,53] A third resolution is to focus on the more severe levels of frailty where definitions converge.[41,54] These solutions may not ensure the clinical applicability of frailty: using frailty as a screening tool (in addition to routine clinical information),[15] using frailty to identify otherwise-recommended interventions, and focusing on severe frailty may not provide added clinical indications beyond current standard age-appropriate care. As per our mapping review, only 2% of appraised articles clearly satisfied the originality criterion.

7.7.3. Recommendations: a robust way forward for frailty research

Ultimately, a frailty diagnosis should alter clinical management by indicating a specific clinical intervention that would not already be indicated. To address barriers and further improve applicability of frailty research that seeks to inform clinical practice, we propose recommendations in Table 3. As a complement to previous reviews and recommendations, we focus specifically on enhancing clinical applicability in the design of studies, interventions, and care pathways, in the reporting and interpretation of results, and when conducting evidence synthesis. A single or consensus definition of frailty may not be achievable nor required considering the multiple functions and disciplines. However, our recommendations highlight that the clinical applicability of any definition of frailty must be tied to a well-defined intervention that improves upon otherwise standard age-related management.

7.7.4. Strengths and limitations

The main strengths of this study include systematic sampling to map and capture a representative sample of the recent frailty research output. We appraised all articles to remove studies where frailty was vaguely defined (e.g., frailty as "old", institutionalized) to select only articles primarily about clinical frailty and maximize potential clinical applicability. Complementing previous reviews, we used a formal framework to classify functions of frailty and appraise how frailty informs or alters current clinical practice. Nonetheless, our study has limitations that deserve mention. First, our appraisal framework and results may be subject to reviewer variability. Although pair of independent reviewers achieved substantial reliability for extraction and appraisal, the classification of articles may differ by discipline and practices, particularly for originality and thus practice-changing status. We have made available our raw results allowing for transparent examination of our appraisal. Second, having mapped published articles only, our appraisal of clinical applicability could be optimistic due to publication bias. Third, we only mapped articles published in 2017 and 2018; research on frailty may have varying trends, and studies published before 2017 could have differing clinical applicability. However, our objective was to examine clinical applicability as a means to advance frailty research, rather than to draw conclusions about specific interventions for frailty. Fourth, because our goal centered on evidence for individuals *already* considered frail, we did not fully appraise articles where frailty was used primarily as an outcome or a confounder; we also only focused on the clinical use of frailty, not on its merit in the research setting which may be substantial. Finally, we did not examine issues related to applying population-level prediction to individuals in clinical practice[55–57].

7.8. CONCLUSION

Although there is rapid growth in the number of publications on frailty, only a minority of this research output provides direct evidence of applicability to practice. The function of frailty, lack of well-defined interventions and identifiable effects, lack of originality, and definition ambiguity hamper its clinical implementation. For frailty to realize its full potential to improve the clinical care and health of older adults, future research on frailty should more deliberately focus on clinical applicability.

7.9. ACKNOWLEDGEMENTS

Author contributions

QDN, EMM, CW designed the study. MFF, HTW, PD, QDN extracted and analyzed the data. QDN drafted the manuscript. All authors contributed significantly to the content, critically reviewed, and approved the final manuscript for publication.

Funding

QDN reports funding from the Canadian Institutes of Health Research and the Fonds de recherche Québec-Santé.

Conflict of interests

None.

7.10. REFERENCES

1. Muscedere J, Andrew MK, Bagshaw SM, et al. Screening for frailty in Canada's health care system: a time for action. *Can J Aging*. 2016;35(3):281-297. doi:10.1017/S0714980816000301

2. Hoogendijk EO, Afilalo J, Ensrud KE, Kowal P, Onder G, Fried LP. Series Frailty 1 Frailty : implications for clinical practice and public health. *Lancet*. 2019;394(10206):1365-1375. doi:10.1016/S0140-6736(19)31786-6

3. Mudge AM, Hubbard RE. Frailty: Mind the gap. *Age Ageing*. 2018;47(4):508-511. doi:10.1093/ageing/afx193

4. Walston JD, Bandeen-Roche K. Frailty: a tale of two concepts. *BMC Med*. 2015;13:185. doi:10.1186/s12916-015-0420-6

5. Mitnitski AB, Mogilner AJ, Rockwood K. Accumulation of deficits as a proxy measure of aging. *Sci World*. 2001;1:323-336. doi:10.1100/tsw.2001.58

6. Searle SD, Mitnitski A, Gahbauer EA, Gill TM, Rockwood K. A standard procedure for creating a frailty index. *BMC Geriatr*. 2008;8(1):24. doi:10.1186/1471-2318-8-24

7. Fried LP, Tangen CM, Walston J, et al. Frailty in older adults: evidence for a phenotype. *J Gerontol A Biol Sci Med Sci*. 2001;56(3):808-813. doi:10.1093/gerona/56.3.M146

8. Bandeen-Roche K, Seplaki CL, Huang J, et al. Frailty in older adults: A nationally representative profile in the United States. *J Gerontol A Biol Sci Med Sci*. 2015;70(11):1427-1434. doi:10.1093/gerona/glv133

9. Buta BJ, Walston JD, Godino JG, et al. Frailty assessment instruments: systematic characterization of the uses and contexts of highly-cited instruments. *Ageing Res Rev*. 2016;26:53-61. doi:10.1016/j.arr.2015.12.003

10. Theou O, Squires E, Mallery K, et al. What do we know about frailty in the acute care setting? A scoping review. *BMC Geriatr*. 2018;18(1). doi:10.1186/s12877-018-0823-2

11. Afilalo J, Alexander KP, Mack MJ, et al. Frailty assessment in the cardiovascular care of older adults. *J Am Coll Cardiol*. 2014;63(8):747-762. doi:10.1016/j.jacc.2013.09.070

12. Turner G, Clegg A. Best practice guidelines for the management of frailty: A British Geriatrics Society, Age UK and Royal College of General Practitioners report. *Age Ageing*. 2014;43(6):744-747. doi:10.1093/ageing/afu138

13. Cesari M, Prince M, Thiyagarajan JA, et al. Frailty: An Emerging Public Health Priority. *J Am Med Dir Assoc*. 2016;17(3):188-192. doi:10.1016/j.jamda.2015.12.016

14. Dent E, Martin FC, Bergman H, Woo J, Romero-ortuno R, Walston JD. Series Frailty 2 Management of frailty: opportunities, challenges, and future directions. *Lancet*. 2019;394(10206):1376-1386. doi:10.1016/S0140-6736(19)31785-4

15. Sourial N, Bergman H, Karunananthan S, et al. Implementing frailty into clinical practice: a cautionary tale. *J Gerontol A Biol Sci Med Sci*. 2013;68(12 A):1505-1511. doi:10.1093/gerona/glt053

16. Hogan DB, Maxwell CJ, Afilalo J, et al. A scoping review of frailty and acute care in middle-aged and older individuals with recommendations for future research approaches to the detection of frailty. *Can Geriatr J*. 2017;20(1):22-37.

17. Karunananthan S, Bergman H. Managing frailty in primary care: evidence gaps cannot be ignored. *Can Med Assoc J*. 2018;190(38):E1122-E1123. doi:10.1503/cmaj.181151

18. Collard RM, Boter H, Schoevers RA, Oude Voshaar RC. Prevalence of frailty in community-dwelling older persons: a systematic review. *J Am Geriatr Soc*. 2012;60(8):1487-1492. doi:10.1111/j.1532-5415.2012.04054.x

19. Dent E, Kowal P, Hoogendijk EO. Frailty measurement in research and clinical practice: a review. *Eur J Intern Med*. 2016;31:3-10. doi:10.1016/j.ejim.2016.03.007

20. de Vries NM, Staal JB, van Ravensberg CD, Hobbelen JSM, Rikkert MGMO, der Sanden MWGN. Outcome instruments to measure frailty: a systematic review. *Ageing Res Rev*. 2011;10(1):104-114. doi:10.1016/j.arr.2010.09.001

21. Sternberg SA, Wershof Schwartz A, Karunananthan S, Bergman H, Mark Clarfield A. The identification of frailty: a systematic literature review. *J Am Geriatr Soc*. 2011;59(11):2129-2138. doi:10.1111/j.1532-5415.2011.03597.x

22. Junius-Walker U, Onder G, Soleymani D, et al. The essence of frailty: a systematic review and qualitative synthesis on frailty concepts and definitions. *Eur J Intern Med*. 2018;56(April):3-10. doi:10.1016/j.ejim.2018.04.023

23. Drubbel I, Numans ME, Kranenburg G, Bleijenberg N, De Wit NJ, Schuurmans MJ. Screening for frailty in primary care: a systematic review of the psychometric properties of the frailty index in community-dwelling older people. *BMC Geriatr*. 2014;14(1):1-13. doi:10.1186/1471-2318-14-27

24. Theou O, Cann L, Blodgett J, Wallace LMK, Brothers TD, Rockwood K. Modifications to the frailty phenotype criteria: systematic review of the current literature and investigation of 262 frailty phenotypes in the survey of health, ageing, and retirement in Europe. *Ageing Res Rev*. 2015;21:78-94. doi:10.1016/j.arr.2015.04.001

25. Aguayo GA, Donneau AF, Vaillant MT, et al. Agreement between 35 published frailty scores in the general population. *Am J Epidemiol*. 2017;186(4):420-434. doi:10.1093/aje/kwx061

26. Hernán MA. Does water kill? A call for less casual causal inferences. *Ann Epidemiol*. 2016;26(10):674-680. doi:10.1016/j.annepidem.2016.08.016

27. Hernán MA, Robins JM. *Causal Inference : What If*. Boca Raton: Chapman & Hall/CRC; 2020.

28. Mason NR, Sox HC, Whitlock EP. A patient-centered approach to comparative effectiveness research focused on older adults: lessons from the Patient-Centered Outcomes Research Institute. *J Am Geriatr Soc*. 2019;67(1):21-28. doi:10.1111/jgs.15655

29. Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. *Heal Inf Libr J*. 2009;26(2):91-108. doi:10.1111/j.1471-1842.2009.00848.x

30. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-174.

31. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. *J Clin Epidemiol*. 1993;46(5):423-429. doi:10.1016/0895-4356(93)90018-V

32. Rockwood K, Song X, MacKnight C. A global clinical measure of fitness and frailty. *Can Med Assoc J*. 2005;173(5):489-495. https://www.cmaj.ca/content/173/5/489.full. Accessed July 13, 2016.

33. Rolfson DB, Majumdar SR, Tsuyuki RT, Tahir A, Rockwood K. Validity and reliability of the Edmonton Frail Scale. *Age Ageing*. 2006;35(5):526-529. doi:10.1093/ageing/afl041

34. Morley JE, Vellas B, Abellan van Kan G, et al. Frailty consensus: a call to action. *J Am Med Dir Assoc*. 2013;14(6):392-397. doi:10.1016/j.jamda.2013.03.022

35. Bergman H, Ferrucci L, Guralnik J, et al. Frailty: an emerging research and clinical paradigm--issues and controversies. *J Gerontol A Biol Sci Med Sci*. 2007;62(7):731-737. doi:62/7/731 [pii]

36. Chiolero A. Counterfactual and interventionist approach to cure risk factor epidemiology. *Int J Epidemiol*. 2016;45(6):2202-2203. doi:10.1093/ije/dyw159

37. Negm AM, Kennedy CC, Thabane L, et al. Management of frailty : a systematic review and network meta-analysis of randomized controlled trials. *J Am Med Dir Assoc*. 2020;20(10):1190-1198. doi:10.1016/j.jamda.2019.08.009

38. Puts MTE, Toubasi S, Andrew MK, et al. Interventions to prevent or reduce the level of frailty in community-dwelling older adults: a scoping review of the literature and international policies. *Age Ageing*. 2017;46(3):383-392. doi:10.1093/ageing/afw247

39. Walston J, Robinson TN, Zieman S, et al. Integrating frailty research into the medical specialties—report from a U13 Conference. *J Am Geriatr Soc*. 2017;65(10):2134-2139. doi:10.1111/jgs.14902

40. Rodríguez-Mañas L, Féart C, Mann G, et al. Searching for an operational definition of frailty: A delphi method based consensus statement. the frailty operative definition-consensus conference project. *J Gerontol A Biol Sci Med Sci*. 2013;68(1):62-67. doi:10.1093/gerona/gls119

41. Xue Q-L, Tian J, Walston JD, Chaves PHM, Newman AB, Bandeen-Roche K. Discrepancy in Frailty Identification: Move Beyond Predictive Validity. *Journals Gerontol Ser A*. 2019;75(2):387-393. doi:10.1093/gerona/glz052

42. Walston J, Buta B, Xue QL. Frailty screening and interventions: considerations for clinical practice. *Clin Geriatr Med*. 2018;34(1):25-38. doi:10.1016/j.cger.2017.09.004

43. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". *Lancet*. 2005;365(9453):82-93. doi:10.1016/S0140-6736(04)17670-8

44. Nguyen QD, Peters E, Wassef A, Desmarais P, Rémillard-Labrosse D, Tremblay-Gravel M. Evolution of age and female representation in the most-cited randomized controlled trials of cardiology of the last 20 years. *Circ Cardiovasc Qual Outcomes*. 2018;11(6):1-11. doi:10.1161/CIRCOUTCOMES.118.004713

45. Odden MC, Peralta CA, Haan MN, et al. Rethinking the Association of High Blood Pressure With Mortality in Elderly Adults. *Arch Intern Med*. 2012;172(15):1046-1058. doi:10.1001/archinternmed.2012.2555

46. Pajewski NM, Williamson JD, Applegate WB, et al. Characterizing Frailty Status in the Systolic Blood Pressure Intervention Trial. *J Gerontol A Biol Sci Med Sci*. 2016;71(5):649-655. doi:10.1093/gerona/glv228

47. Sanders NA, Supiano MA, Lewis EF, et al. The frailty syndrome and outcomes in the TOPCAT trial. *Eur J Heart Fail*. 2018;20(11):1570-1577. doi:10.1002/ejhf.1308

48. Meneilly GS, Knip A, Miller DB, Sherifali D, Tessier D, Zahedi A. Diabetes in Older People. *Can J Diabetes*. 2018;42:S283-S295. doi:10.1016/j.jcjd.2017.10.021

49. Grieve AP. The number needed to treat: a useful clinical measure or a case of the Emperor's new clothes? *Pharm Stat*. 2003;2(2):87-102. doi:10.1002/pst.33

50. Bouillon K, Kivimaki M, Hamer M, et al. Measures of frailty in population-based studies: an overview. *BMC Geriatr*. 2013;13(1):64. doi:10.1186/1471-2318-13-64

51. Pialoux T, Goyard J, Lesourd B. Screening tools for frailty in primary health care: A systematic review. *Geriatr Gerontol Int*. 2012;12(2):189-197. doi:10.1111/j.1447-0594.2011.00797.x

52. Morley JE, Vellas B, van Kan GA, et al. Frailty Consensus: A Call to Action. *J Am Med Dir Assoc*. 2013;14(6):392-397. doi:10.1016/j.jamda.2013.03.022

53. Dent E, Lien C, Lim WS, et al. The Asia-Pacific clinical practice guidelines for the management of frailty. *J Am Med Dir Assoc*. 2017;18(7):564-575. doi:10.1016/j.jamda.2017.04.018

54. Rockwood K, Andrew M, Mitnitski A. A comparison of two approaches to measuring frailty in elderly people. *J Gerontol A Biol Sci Med Sci*. 2007;62(7):738-743. doi:10.1093/gerona/62.7.738

55. Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *Am J Epidemiol*. 2004;159(9):882-890. doi:10.1093/aje/kwh101

56. Henderson R, Keiding N. Individual survival time prediction using statistical models. *J Med Ethics*. 2005;31(12):703-706. doi:10.1136/jme.2005.012427

57. Pijpers E, Ferreira I, Stehouwer CDA, Nieuwenhuijzen Kruseman AC. The frailty dilemma. Review of the predictive accuracy of major frailty scores. *Eur J Intern Med*. 2012;23(2):118-123. doi:10.1016/j.ejim.2011.09.003

7.11. TABLES

Table 7.1. Study and Population Characteristics, Interventions, and Primary Outcomes in a Random Sample of 150 articles Published on Frailty in 2017-2018

Journal discipline (n=150)	**n (%)**	**Year of publication (n=150)**	**n (%)**
Aging and geriatrics	47 (31)	2017	69 (46)
Medical subspecialties	30 (20)	2018	81 (54)
Surgery and anesthesia	25 (17)	**Study design (n=150)**	
Epidemiology, HSR, public health	11 (7)	Cohort	137 (91)
General and internal medicine	9 (6)	Trial	13 (9)
Psychiatry	9 (6)	*Randomized controlled trial*	*10 (7)*
Allied health professions	6 (4)		
Other	13 (9)		
Sample size, median (IQR)	476 (151-1891)	**Interventions in trials (n=13)**	**n**
Age, average (SD)	72.6 (9.8)	Geriatric prevention and multidomain	6
Percentage of women (SD)	53 (19)	intervention	
Study enrollment setting, n (%)		Exercise training	3
Community-dwelling	53 (35)	Perioperative multifactorial intervention	2
Institution	2 (1)	Shared decision-making training	1
Mixed, hospital-based, unspecified	95 (63)	Bladder and exercise training	1
Population when hospital-based enrollment, n=82	**n (%)**	**Primary outcomes (n=150)**	**n (%)**
Surgery	34 (41)	Clinical	106 (71)
General or GI	*12 (15)*	*Mortality*	*49 (33)*
Orthopedic	*7 (9)*	*Clinical condition or marker*	*17 (11)*
Cardiovascular	*5 (6)*	*Morbidity*	*13 (9)*
Urologic	*4 (5)*	*Psychiatric symptoms*	*7 (4)*
Mixed or other	*6 (7)*	*Medication appropriateness*	*6 (4)*
Cardiology	15 (18)	*Cognition*	*5 (3)*
Heart failure	*5 (6)*	*Frailty*	*5 (3)*
TAVI	*4 (5)*	*Fall, fracture, fear of falling*	*4 (3)*
ASCVD	*3 (4)*	Health care service	26 (17)
Hypertension	*2 (2)*	*Health care usage*	*10 (7)*
Other	*2 (2)*	*Discharge disposition*	*6 (4)*
		Clinical process improvement	*2 (1)*

Internal or geriatric medicine	7 (9)	Patient-centered	18 (12)
Oncology	6 (7)	*Function*	*10 (7)*
Intensive care	5 (6)	*Quality of life*	*8 (5)*
Nephrology	4 (5)	Other	8 (5)
Acute care hospital	3 (4)		
Emergency medicine	2 (2)		
Respirology	2 (2)		
Other	4 (5)		

Notes. ASCVD = atherosclerotic cardiovascular disease, GI = Gastrointestinal, HSR = health services research, TAVI = transcatheter aortic valve implantation

Table 7.2. Definitions and Cut-Offs for Frailty

Frailty definition (n=161), 38 definitions*	Count, n (%)	Cut-off for frailty, mean (SD)	Cut-off for prefrailty, mean (SD)
Fried frailty phenotype	45 (28)		
Fried	39 (24)	3	2
Fried modifications	6 (4)	Ad hoc	Ad hoc
Frailty index-derived	35 (22)	0.28 (0.08) Range 0.18-0.41	0.15 (0.07) Range 0.10-0.26
Frailty index	21 (13)	0.29 (0.06) Range 0.21-0.41	0.15 (0.07) Range 0.10-0.26
Modified frailty index 11	11 (7)	0.23 (0.08) Range 0.18-0.36	0.09 (-)
Modified frailty index 5	2 (1)	0.40	0.25
Electronic frailty index	1 (1)	0.24	0.12
Clinical frailty scale	16 (10)	4.9 (0.9) Range 3-6	
		Cut-off for frailty	
Grip strength and/or gait speed	11 (7)	Median by age and/or sex Fried criteria by sex, BMI 26 kg men, 18 kg women Subjective < 0.22 bar < 1.0 m/s < 0.8 m/s < 0.6 m/s	
Edmonton frail scale	5 (3)	8, 8, 10, 12	
Groningen frailty indicator	4 (2)	4, 4, 4, 5	
CGA-based	3 (2)	Ad hoc	
FRAIL scale	3 (2)	3	
Kihon checklist	3 (2)	8	
Tillburg frailty indicator	3 (2)	5	
5MWT, ADL, Braden, CFS+MoCA, impaired cognition + dynapenia, Elderly mobility scale, Frailty score - Guilley, Frailty score - Porock, Frailty score - Robinson, Frailty score, Functional independence measure, G8, Gérontopôle frailty screening tool, HFRS, ISAR, ISAR-HP, John Hopkins Adjusted Clinical Groups, Kaigo-Yobo Check-List, Katz, Morse, Risk analysis index, Schoenenberger	1 or 2 (0.6 or 1.2) Total: 33 (20)	-	

frailty index, SHARE-FI, social frailty, SOF, temporal muscle thickness, VES-13, VMS	

Notes. * Supplementary Table 7.4 lists main references for frailty definitions. 5MWT = 5-minute walk test, CFS = Clinical frailty scale, MoCA = Montreal Cognitive Assessment, HFRS = Hospital frailty risk score, ISAR = Identification of seniors at risk, ISAR-HP = Identification of seniors at risk – hospitalized patients, SOF = Study of Osteoporotic Fractures, VES-13 = Vulnerable elders survey-13, VMS = VeiligheidsManagementSysteem.

Table 7.3. Recommendations to address barriers and improve applicability of frailty research seeking to inform clinical practice

Recommendations	Design of studies, interventions, care pathways	Reporting and interpretation	Evidence synthesis
1. The definition of frailty used should be precisely described: a. In addition to the conceptual framework used, criteria and cut-offs for each component or item should be explicitly reported to allow replication and appraisal of generalizability b. If distributional cut-offs (e.g., lowest quintile) are used, absolute values should be reported	X	X	X
2. In establishing a precise definition of frailty, additionally consider context-of-use and purpose-of-use, namely: a. The function of frailty: whether it is a predictor for which a frailty-specific intervention is required (etiologic), whether it is an effect modifier to assess whether an otherwise-indicated intervention would provide net benefit, or whether it is a selection criterion for an intervention. b. If frailty is a selection criterion, consider the potential intervention that would follow the identification of frailty. Choose frailty definition and include domains likely to select subjects most likely to benefit from ensuing frailty interventions. Aim to use frailty as a means to indicate interventions that are not already components of standard age-appropriate care. c. Consider the setting (community-dwelling or acute care), discipline, duration (short-term or long-term), and thus feasibility.	X		

3. In the context of observational studies reporting on a non-frailty specific intervention (e.g., surgery), provide outcomes for individuals considered frail on the absolute scale to allow for comparison to projected outcome of forgoing the intervention.		X	
4. Shift the focus from association or predictive frailty studies to conduct intervention trials where frailty is a selection criterion or an effect measure modifier (for sufficiently powered trials).	X		
5. Avoid considering specific operationalizations of frailty as *a priori* interchangeable.		X	X
6. Statement of clinical applicability should be reserved for primary and substantiated evidence.		X	X

7.12. FIGURES

Figure 7.1. Flowchart for selection, classification, and full appraisal of articles

Figure 7.1A. Flowchart for selection and classification of articles

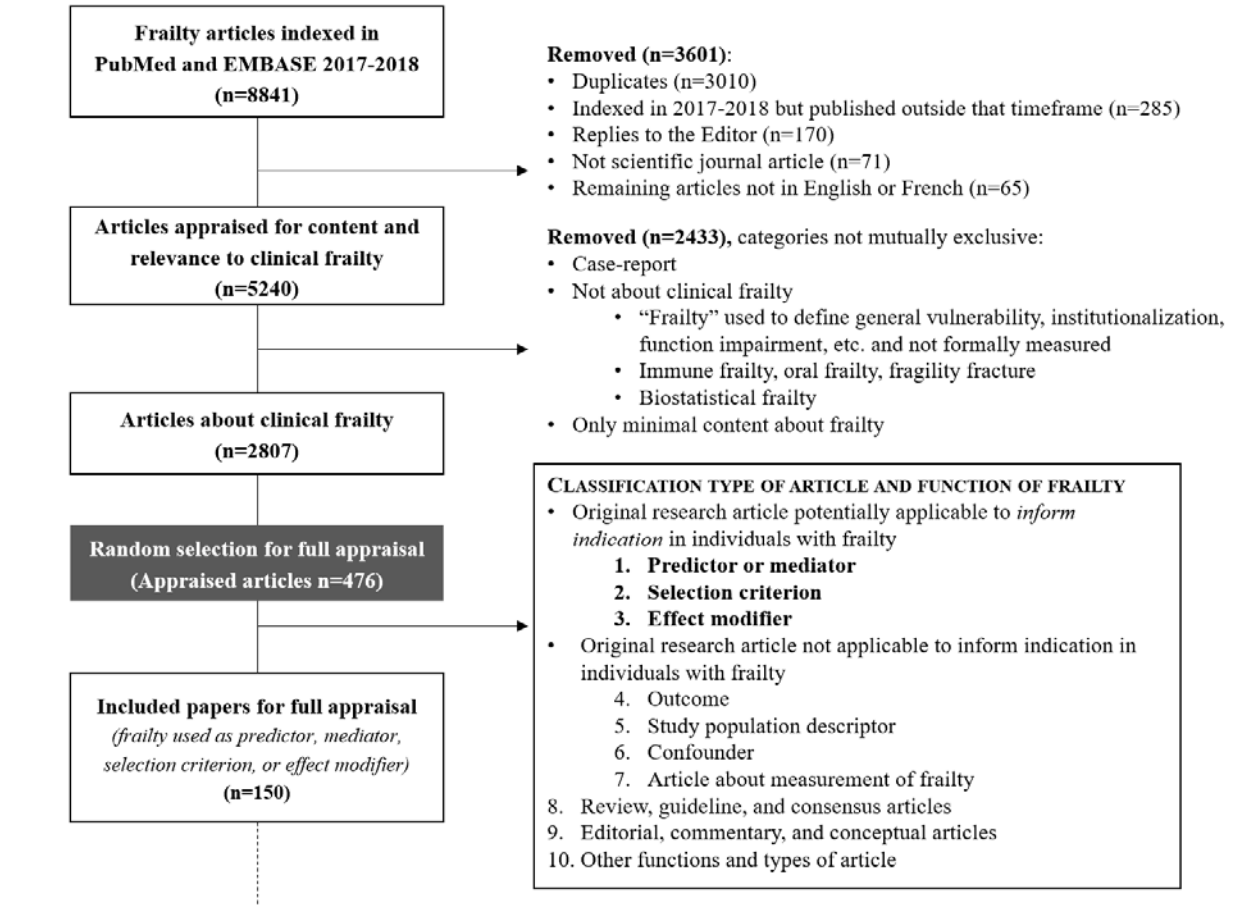


Figure 7.1B. Flowchart for full appraisal of articles

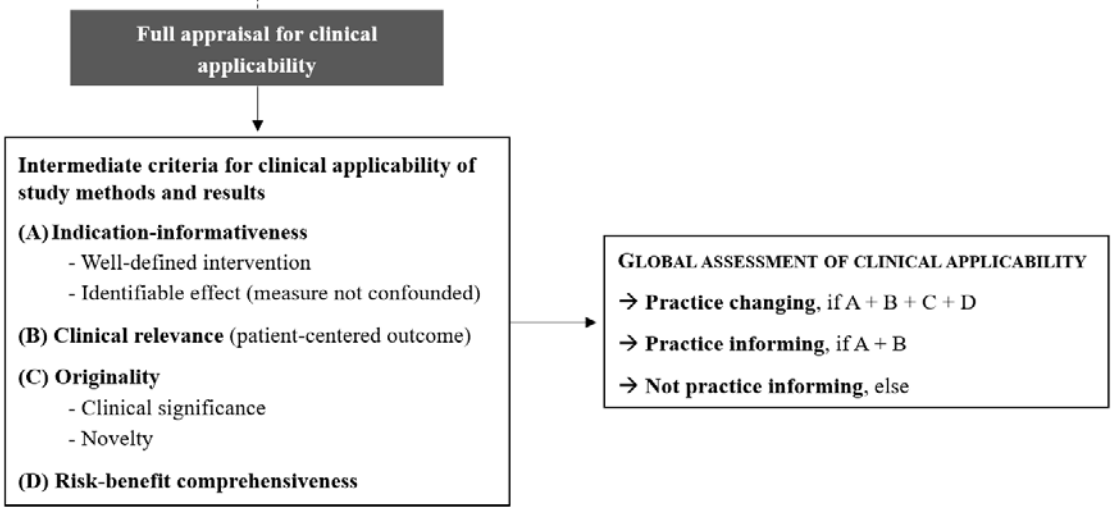


Figure 7.2. Distribution of articles on frailty by function and article type, and clinical applicability

Figure 7.2A. Distribution of articles on frailty by function and article type, from a random sample of 476 articles published in 2017-2018

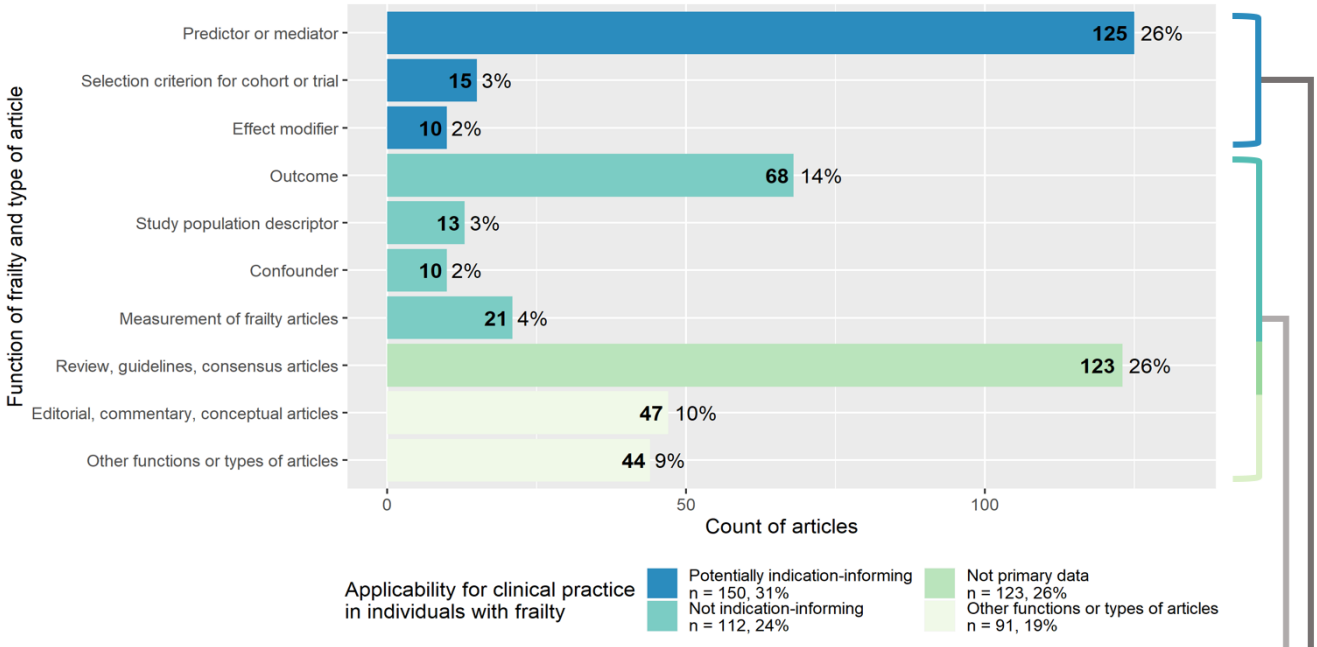


Figure 7.2B. Clinical applicability of articles on frailty by function and study design

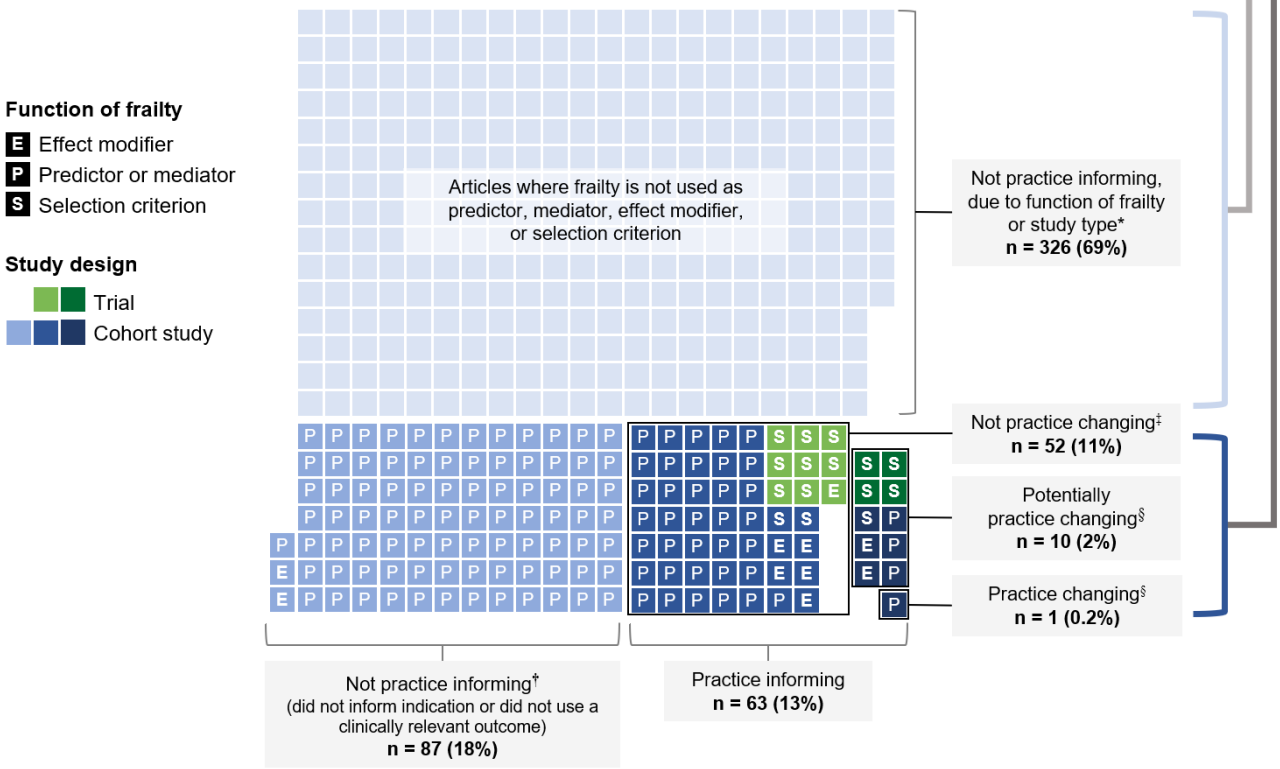


Figure 7.2 Legend.
* 326 articles were not practice informing because of the function of frailty or study type. † 87 articles were classified as not practice informing because they (i) did not inform clinical indication (no well-defined intervention for frailty, the effect of that intervention was not identifiable, or there was no report of absolute outcome results for an intervention on individuals with frailty), or (ii) did not use a clinical relevant outcome. ‡ Of the 52 articles classified as practice informing but not practice changing, all 36 articles where frailty was a predictor or mediator [P] were informative by reporting absolute outcome results of an intervention on individuals with frailty, and 9 trials did not report results considered original. § By (potentially) fulfilling all 4 intermediate criteria (indication-informativeness, clinical relevance, originality, and risk-benefit comprehensiveness), 10 articles were classified as potentially practice changing and 1 as practice changing.

See Methods and Supplementary Methods 2 for definitions and descriptions of clinical applicability.

7.13. SUPPLEMENTAL MATERIAL

Supplementary Methods 1. Search Strategy and Frailty Functions Description

SEARCH STRATEGY FOR PUBMED MEDLINE

((frailty[mesh]) OR (frail elderly[mesh]) OR (frail*[tiab])) AND ("french"[Language] OR "english"[Language])

Filters: Publication date from 2017/01/01 to 2018/12/31

SEARCH STRATEGY FOR OVID-EMBASE

#	Searches
1	frail elderly/ or frail*.mp. or frailty/
2	limit 1 to yr="2017 - 2018"
3	limit 2 to embase
4	limit 3 to (english or french)

Search conducted on April 8, 2019

FRAILTY FUNCTION DESCRIPTIONS

Function	Description	Example
Predictor	Frailty is a predictor/independent variable of interest in relation to an outcome.	Study investigating the association of frailty with mortality
Mediator	Frailty is investigated as an *intermediate* in the association between another variable and outcome.	Study investigating whether frailty in the intermediate/mechanism between the association of low socioeconomic status and subjective well-being
Selection criterion	Frailty is a primary selection/eligibility criterion for a trial/experimental study.	Experimental study of a multidimensional intervention conducted in individuals with frailty
Effect modifier	The association between another variable and outcome is examined by *levels* of frailty (by subgroups of individuals with frailty vs. without frailty)	In a trial: a study or analysis investigating whether intensive blood pressure management reduces mortality, *differentially* in those with frailty compared to those without frailty. In a cohort study: a study or analysis investigating whether influenza vaccination has a *different* effect on immunity in individuals with frailty compared to those without frailty.
Study population descriptor	The distribution of frailty is reported (typically in Table 1) to better characterize the study population.	Frailty is reported along with sociodemographic and chronic condition variables in cohort study or trial.

Confounder	Frailty is used as a stratifying variable or as a covariate to control/adjust for confounding.	Vaccine effectiveness is analyzed by levels of frailty (stratification); frailty is included in a regression model with vaccination as independent variable and mortality as outcome.

Supplementary Methods 2. Definitions, Matrix for Determination, and Descriptions of Clinical Applicability

DEFINITIONS OF INTERMEDIATE CRITERIA

Intermediate Criteria of Clinical Applicability	**Definition**
(A) Indication-informativeness	Study methods provide clinicians with evidence to determine a clinical indication, (intervention or potential intervention) in individuals with frailty. Informativeness for a clinical indication for individuals with frailty requires a well-defined intervention, whose effect can be identified from the study results, i.e.: 1. A trial of an intervention in individuals with frailty; **OR** 2. An observational study where: A) A well-defined intervention for individuals with frailty exists, **AND** B) That the effect of this well-defined intervention be correctly estimated (no confounding – independent effects of the intervention); **OR** 3. An observational study where there is an intervention on individuals with frailty and where *absolute* results for outcomes are explicitly reported. (*informativeness for the outcome of an intervention* criterion – allows contrast between intervention in individuals with frailty and envisioned natural history under no intervention in individuals with frailty).
(B) Clinical relevance	Primary outcome of the study is clinically relevant, i.e., the outcome is patient centered.
(C) Originality	Study results achieve statistical and clinical significance (e.g., a relevant magnitude of effect); **AND** Study results are novel when compared to current evidence base and practice (novelty).
(D) Risk-benefit comprehensiveness	Overall benefits of changing an indication (either the intervention or the population of individuals in which the intervention is indicated) outweigh the risks.

MATRIX FOR DETERMINATION OF CLINICAL APPLICABILITY BASED ON INTERMEDIATE CRITERIA

		INTERMEDIATE CRITERIA FOR CLINICAL APPLICABILITY			
		(A) Indication-informativeness	**(B) Clinical relevance**	**(C) Originality**	**(D) Risk-benefit comprehensiveness**
CLINICAL APPLICABILITY	**Practice changing**	Yes	Yes	Yes	Yes
	Practice informing	Yes	Yes	No	No
	Not practice informing	No	No	No	No

DESCRIPTIONS OF CLINICAL APPLICABILITY

Clinical Applicability	Description
Practice changing	**Requires intermediate criteria A + B + C + D** Articles that are *practice changing* provide evidence that alters current clinical practice. As such, they *inform a clinical indication (A)*, report a *clinically relevant outcome (B)*, have *original results (C)* and a positive *comprehensive risk-benefit assessment (D)*.
Practice informing	**Requires intermediate criteria A + B** Articles that are *practice informing* provide evidence that applies to clinical practice. Methods and results from these studies *inform a clinical indication (A)* and have *clinically relevant outcome (B)* for individuals with frailty. They inform practice by supplementing the current evidence but do not change current accepted clinical practice (not *original (C)* or *no positive risk-benefit assessment (D)* of altering practice).
Not practice informing	**Does not minimally meet intermediate criteria A + B** Articles that are *not practice informing* do not provide evidence that does is directly applicable to clinical practice for older adults with frailty. They belong to the research domain (in contrast to the clinical practice domain).

Supplementary Table 7.1. Agreement, Kappa and Prevalence-Adjusted Bias-Adjusted Kappa by Appraisal Step, and Overall

Appraisal step	Agreement, %	Kappa, unweighted (95% CI)	Prevalence-adjusted bias-adjusted kappa (95% CI)
Well-defined intervention and effect identifiable	95	-0.03 (-0.16, 0.11)	0.85 (0.73, 0.93)
Informativeness for the outcome of an intervention	88	0.72 (0.55, 0.89)	0.76 (0.62, 0.86)
Appropriate indication-informing methods	88	0.74 (0.59, 0.90)	0.76 (0.63, 0.85)
Originality	82	0.39 (0.12, 0.65)	0.64 (0.38, 0.82)
Clinical relevance of outcome	98	0.88 (0.62, 1.00)	0.96 (0.81, 1.00)
Patient-centered outcome	93	0.71 (0.45, 0.97)	0.86 (0.66, 0.96)
Overall risk-benefit comprehensiveness	83	-*	0.40 (-0.30, 0.87)
Global assessment to inform practice	89	0.76 (0.60, 0.92)	0.77 (0.65, 0.87)
Global assessment to change practice	91	0.37 (0.21, 0.53)	0.81 (0.70, 0.90)
Statement of clinical applicability	79	0.52 (0.37, 0.68)	0.57 (0.42, 0.70)
Overall	**87**	**0.74 (0.68, 0.81)**	**0.76 (0.72, 0.80)**

Notes. * Calculated kappa was 0 (0, 0) because proportion of observed = 0.7, proportion expected "yes" = 0.7, and proportion expected "no" = 0.

Supplementary Table 7.2. Intermediate Criteria and Global Appraisal Results for 150 Articles Fully Appraised

Criteria	Number of articles appraised, n*	Yes, n (%)	No, n (%)	Uncertain, n (%)
(A) Indication-informativeness	150	60 (40)	87 (58)	3 (2)
Well-defined intervention and identifiable effect	137	5 (4)	131 (96)	1 (1)
Informativeness for outcome in individuals with frailty	140	51 (36)	87 (62)	2 (1)
(B) Patient-centered outcome	63	52 (83)	6 (10)	5 (8)
(C) Originality	63	1 (2)	52 (83)	10 (16)
(D) Risk-benefit comprehensiveness	10	6 (60)	0 (0)	4 (40)
Global assessment to inform practice	150	60 (40)	90 (60)	0 (0)
Global assessment to change practice	150	1 (1)	137 (91)	12 (8)

Notes. * Intermediate criteria were appraised hierarchically. Indication-informativeness was determined by frailty function (some articles satisfied *Indication-informativeness* due to their function as effect modifier or selection criterion), well-defined intervention/identifiable effect, and informativeness for outcome. Originality and patient-centered outcome were only appraised in the 63 articles satisfying criterion A. Risk-benefit comprehensiveness was only appraised in the 10 articles satisfying criterion A, B, and C.

Supplementary Table 7.3. Classification of Interventions for Frailty Reported in 62 of the 150 Appraised Articles

Exercise and multidomain preventive and lifestyle intervention, n=17	General, n=10 Exercise programs Physical exercise and nutritional programs Standardized strength training program in combination with nutritional support Exercise, nutrition, and pharmacological agents Multifactorial intervention: activities focusing on physical exercise, the Mediterranean diet, assessment of inadequate prescription in polypharmacy patients and social assessment High and low-level care including education, problem solving therapy, exercise sessions Multimodal lifestyle interventions targeted at common psychosocial and biological factors Preventive interventions Preventive action Interventions to reduce the impact of stress Frailty-targeted, n=7 Health care interventions focused on reducing frailty Early frailty prevention programs targeted to maintain muscle strength and gait speed with physical exercise, adequate nutrition, and fall prevention Interventions to modify frailty and ameliorate its effects, optimising care, and planning interventions, risk stratification, clinical guidelines, health-care delivery, and design and planning of interventions Strategies targeting the primary and secondary prevention of age-related frailty Guide treatment choices, integrated care, targeting both frailty and depression, assistance in daily living, vitamin D supplementation, protein enriched diet, and improving exercise frequency Adequate nutrition and optimization of physical activity, multi-faceted approach to therapy in high-risk patients, targeting the potentially modifiable factors of depressive symptoms, cognitive impairment, and physical frailty. Risk stratification, strategies specifically targeted to prevent the process of deconditioning or to slow down the transition to dependence, rational health care resources management
General tailoring and targeting interventions, n=11	Targeted intervention Tailor risk assessment and interventions Better tailored interventions Optimise therapeutic approach Individualised approach Aggressive interventions, diagnostic efforts, or specific therapeutic interventions Deciding on therapeutic approach, planning health resources Personalization of therapy, adjust treatment protocols Intervention strategies, tailored frailty care plans based on the whole individual, positively influence care plan decisions appropriate for frail patients Personalize treatment decisions for older adults, targeted frailty interventions Comprehensive intervention strategies
Context or condition-specific tailoring and targeting interventions, n=6	Cardiovascular, n=3 Radial access, properly dose-adjusted anticoagulant therapies Guideline-recommended medications for ACS Strategy for surgical AVR or TAVR, strategic decision-making ICU, n=1 Manage the factors associated with ICU mortality, decrease aggressive interventions, more effective planning and decisions about the life-sustaining treatments or palliative care

	Surgery and specific, n=2 Tailor the invasiveness of the surgery Intervention for prevention or treatment of post-operative delirium
Comprehensive surgery management including Shared decision-making, n=8	Preoperative risk stratification, more careful monitoring, available guidelines, such as those provided by the American Geriatrics Society/American College of Surgeons, enhancement their physiologic status in a reasonable timeframe before surgery Informed consent for surgical procedures, choice of anaesthesia, pain management, rehabilitation post-surgery, early mobilisation, early detection and management of geriatric syndromes, adopt more conservative management Risk stratification that may guide surgical and medical decision-making, shared decision-making between patient and provider, enhanced preoperative optimization of medical comorbidities, more targeted pre- and post-discharge interventions, such as earlier and more frequent follow-ups via nursing telephone calls or outpatient clinic visits Guide perioperative care in an interdisciplinary team setting, specialty teams, proactive patient counselling, interventions to decrease the associated high-cost burden to the health care system Guide decision-making, aid in the informed consent process, efficient allocation of hospital resources, opportunities for early intervention Crucial conversations, informed decision-making, allocating greater hospital resources, preoperative counseling, and postoperative planning Early focused intervention, tailoring the proper management, clinical decision-making, proper resource mobilization for preventive interventions, risk stratification and to standardize the long-term management, guide the involvement of multidisciplinary geriatric team, early administration of therapeutic and preventive interventions Modification of risk factors to decrease mortality; interventions, such as flagging frail patients for administrative review, identifying patients at risk for F2R, management of complications to increase rescue in vulnerable populations, and prehabilitation; select, optimize, and modify care patterns; modification of the treatment plan and postoperative management; discussion with the care team, potential modification of the surgical plan, and formal preoperative palliative care consultation; early escalation of care, improved situational awareness, and centralization of care at high-volume centers for high-risk patients
Geriatric, multidisciplinary, and adapted care n=11	Geriatric and multidisciplinary, n=5 Interventions such as multidisciplinary consultation Intensive specialist follow-up targeted to frail older adults, mainly by Geriatricians Comprehensive geriatric assessment Building frailty clinics for an in-depth assessment and incorporating physical and cognitive exercise, social support, and nutrition Maximise accessibility to geriatric ward Adapted care pathways, n=2 Care pathways Fastrack admission, faster activation of multidisciplinary team Rehabilitation, n=4 Design different rehabilitation regimens Additional resources before, rehabilitation facility placement postoperatively Rehabilitation Health care resources, physical exercise
Palliative care and alternatives to intervention, n=5	Palliative treatment Palliative care intervention Shared decision-making, preferential palliative care More aggressive or palliative approach Potential alternatives to the surgery
Appropriateness of medications, n=4	Strategies to minimize inappropriate medication Monitoring of polypharmacy, appropriate prescription and adherence tools and a tight medicines control Check whether all medications are necessary, evidence-based and appropriate, and whether there are relevant interactions, medication reviews Improving the therapeutic management, adapted and person-tailored interventions, conservative approach avoiding drastic pharmacological interventions

Supplementary Table 7.4. Main reference for frailty definitions identified

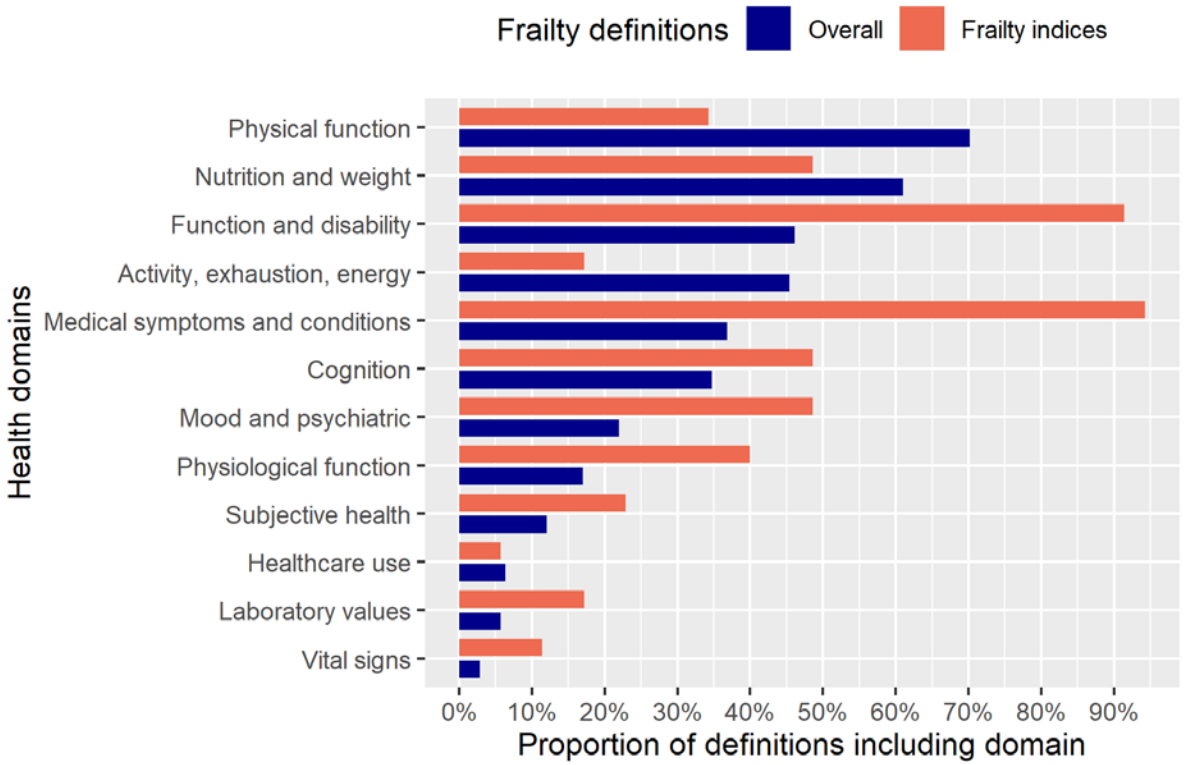
Frailty definition	Main reference
5-minute walk test	-
Activities of daily living	Katz, S, Ford, AB, Moskowitz, RW Jackson, BA, Jaffe, MW. Studies of illness in the aged: The Index of ADL: A standardized measure of biological and psychosocial function. JAMA. 1963;185(12), 914-919.
Braden scale	Bergstrom N, Braden BJ, Laguzza A, Holman V. The Braden Scale for Predicting Pressure Sore Risk. Nurs Res. 1987;36(4):205–210.
Clinical frailty scale	Rockwood K, Song X, MacKnight C. A global clinical measure of fitness and frailty. Can Med Assoc J. 2005;173(5):489-495.
Clinical frailty scale +MoCA	Rockwood K, Song X, MacKnight C. A global clinical measure of fitness and frailty. Can Med Assoc J. 2005;173(5):489-495. Nasreddine ZS, Phillips NA, Bédirian V, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc. 2005;53(4):695–699.
Comprehensive Geriatric Assessment-based	Balducci L, ExtermannM. Management of cancer in the older person: a practical approach. Oncologist. 2000;5(3):224–237.
Edmonton frail scale	Rolfson DB, Majumdar SR, Tsuyuki RT, Tahir A, Rockwood K. Validity and reliability of the Edmonton Frail Scale. Age Ageing. 2006;35(5):526-529.
Elderly mobility scale	Smith R. Validation and reliability of the Elderly Mobility Scale. Physiotherapy 1994;80:744-7.
FRAIL scale	Abellan van Kan G, Rolland YM, Morley JE, et al. Frailty: toward a clinical definition. J Am Med Dir Assoc. 2008;9(2): 71–72.
Frailty index	Mitnitski AB, Mogilner AJ, Rockwood K. Accumulation of deficits as a proxy measure of aging. Sci World. 2001;1:323-336.
Modified frailty index 11	Velanovich V, Antoine H, Swartz A, Peters D, Rubinfeld I. Accumulating deficits model of frailty and postoperative mortality and morbidity: its application to a national database. J Surg Res. 2013;183(1):104-110.
Modified frailty index 5	Subramaniam S, Aalberg JJ, Soriano RP, Divino CM. New 5-Factor Modified Frailty Index Using American College of Surgeons NSQIP Data. J Am Coll Surg. 2018;226(2):173-181.e8.
Frailty score - Guilley	Guilley E, Ghisletta P, Armi F, Berchtold A, Lalive d'Epinay C, Michel JP, de Ribaupierre A. Dynamics of frailty and ADL dependence in a five-year longitudinal study of octogenarians. Res Aging. 2008;30:299–317

Frailty definition	Main reference
Frailty score - Porock	Porock D, Parker-Oliver D, Petroski GF, Rantz M. The MDS Mortality Risk Index: The evolution of a method for predicting 6-month mortality in nursing home residents. BMCRes Notes. 2010;3:200-207.
Frailty score - Robinson	Robinson TN, Wu DS, Pointer L, et al. Simple frailty score predicts postoperative complications across surgical specialties. Am J Surg 2013;206:544–50.
Frailty score - Dodson	Dodson JA, Hochman JS, Roe MT, et al. The Association of Frailty With In-Hospital Bleeding Among Older Adults With Acute Myocardial Infarction: Insights From the ACTION Registry. JACC Cardiovasc Interv. 2018;11(22):2287–2296.
Fried frailty phenotype	Fried LP, Tangen CM, Walston J, et al. Frailty in older adults: evidence for a phenotype. J Gerontol A Biol Sci Med Sci. 2001;56(3):808-813.
Functional independence measure	Heinemann AW, Linacre JM, Wright BD, Hamilton BB, Granger C. Relationships between impairment and physical disability as measured by the functional independence measure. Arch Phys Med Rehabil. 1993;74:566-573.
G8	Bellera CA, Rainfray M, Mathoulin-Pélissier S, et al. Screening older cancer patients: first evaluation of the G-8 geriatric screening tool. Ann Oncol. 2012;23(8):2166–2172.
Gérontopôle frailty screening tool	Vellas B, Balardy L, Gillette-Guyonnet S, et al. Looking for frailty in community-dwelling older persons: the Gérontopôle Frailty Screening Tool (GFST). J Nutr Health Aging. 2013;17(7):629–631.
Grip strength and/or gait speed	-
Groningen frailty indicator	Stevernik N, Slaets JPL, Schuurmans H, et al. Measuring frailty: development and testing the GFI (Groningen Frailty Indicator). Gerontologist. 2001;41(special issue 1):236–237.
Hospital frailty risk score	Gilbert T, Neuburger J, Kraindler J, et al. Development and validation of a Hospital Frailty Risk Score focusing on older people in acute care settings using electronic hospital records: an observational study. Lancet. 2018;391(10132):1775-1782.
Impaired cognition + dynapedia	-
Identification of seniors at risk	McCusker J, Bellavance F, Cardin S, et al. Detection of older people at increased risk of adverse health outcomes after an emergency visit: the ISAR screening tool. J Am Geriatr Soc 1999;47:1229-37.
Identification of seniors at risk – hospitalized patients	Hoogerduijn JG, Buurman BM, Korevaar JC, Grobbee DE, de Rooij SE, Schuurmans MJ. The prediction of functional decline in older hospitalised patients. Age Ageing. 2012;41(3):381-387.

Frailty definition	Main reference
John Hopkins Adjusted Clinical Groups	Lieberman R, Abrams C, Weiner J. Development and Evaluation of the Johns Hopkins University Risk Adjustment Models for Medicare+Choice Plan Payment. Baltimore, MD: Centers for Medicare and Medicaid Services, U.S. Department of Health and Human Services; 2003.
Katz	Katz, S, Ford, AB, Moskowitz, RW Jackson, BA, Jaffe, MW. Studies of illness in the aged: The Index of ADL: A standardized measure of biological and psychosocial function. JAMA. 1963;185(12), 914-919.
Kaigo-Yobo Check-List	Shinkai S, Watanabe N, Yoshida H et al. Research on screening for frailty: development of “the Kaigo-Yobo Checklist”. Nihon Koshu Eisei Zasshi 2010; 57:345–354.
Kihon checklist	Sewo Sampaio PY, Sampaio RA, Yamada M, Arai H. Systematic review of the Kihon Checklist: Is it a reliable assessment of frailty? Geriatr Gerontol Int 2016;16:893–902.
Morse	McCollam ME. Evaluation and implementation of a research-based falls assessment innovation. Nurs Clin North Am. 1995;30(3):507–514.
Risk analysis index	Johnson MS, Bailey TL, Schmid KK, Lydiatt WM, Johanning JM (2014) A frailty index identifies patients at high risk of mortality after tracheostomy. Otolaryngol Head Neck Surg 150(4):568–573.
Schoenenberger frailty index	Schoenenberger AW, Stortecky S, Neumann S, et al. Predictors of functional decline in elderly patients undergoing transcatheter aortic valve implantation (TAVI). Eur Heart J 2013;34:684–92.
SHARE-FI	Romero-Ortuno R, Walsh CD, Lawlor BA, Kenny RA. A frailty instrument for primary care: findings from the Survey of Health, Ageing and Retirement in Europe (SHARE). BMC Geriatr. 2010;10:57.
Social frailty	Makizako H, Shimada H, Tsutsumimoto K, et al. Social Frailty in Community-Dwelling Older Adults as a Risk Factor for Disability. J Am Med Dir Assoc. 2015;16(11):1003.
Study of Osteoporotic Fractures	Ensrud KE, Ewing SK, Taylor BC, et al. Frailty and risk of falls, fracture, and mortality in older women: the Study of Osteoporotic Fractures. J Gerontol A Biol Sci Med Sci. 2007; 62(7):744–751.
Temporal muscle thickness	Ranganathan K, Terjimanian M, Lisiecki J, Rinkinen J, Mukkamala A, Brownley C et al. Temporalis muscle morphomics: the psoas of the craniofacial skeleton. J Surg Res. 2014;186:246–252
Tillburg frailty indicator	Gobbens RJJ, van Assen MALM, Luijkx KG, Wijnen-Sponselee MT, Schols JMGA. The Tilburg frailty indicator: Psychometric properties. J Am Med Dir Assoc. 2010;11(5):344-355.
Vulnerable elders survey-13	Saliba D, Elliott M, Rubenstein LZ, et al. The Vulnerable Elders Survey: a tool for identifying vulnerable older people in the community. J Am Geriatr Soc. 2001;49(12):1691–1699.

Frailty definition	Main reference
VeiligheidsManagementSysteem	VMS. Praktijkgids kwetsbare oudere. Available from: http://www. vmszorg.nl/_library/5540/web_2009.0104_praktijkgids_kwetsbare_ ouderen.pdf

Supplementary Figure 7.1. Proportion of Health Domains Included in Frailty Definitions, Overall and for Frailty Indices



Chapter 8. Discussion and conclusion

8.1 Summary of findings

In a very general sense, frailty aims to categorize individuals. Frailty lives in the space of clinical epidemiology where a critical purpose of categorization is to allow eventual implementation in clinical practice to improve health outcomes. Findings presented in this dissertation provide insights into whether frailty, as an instance of an aged-related nosological construct, may achieve or has achieved this goal.

In Manuscript 1, I explored the distribution and patterns of variability in health characteristics as a function of chronological age and demonstrated that heterogeneity generally increased with age, albeit not uniformly. Of 34 health characteristics, 17 showed increased heterogeneity, eight decreased heterogeneity, and nine showed no association with age. Associations were nonlinear for most domains and non-monotonic for some. A close examination of heterogeneity showed its intertwining with measurements properties: associations between heterogeneity and chronological age may be due to relationships between the mean and variance of distributions of variables and to the clinical scaling of measurement instruments. Results from Manuscript 1 strongly suggest that there is a basis for the past and continued development of age-relate constructs such as frailty.

In Manuscript 2, I examined the ability of deficit-accumulation frailty indices, as an influential exemplar of age-related constructs, to achieve successful *clinical* categorization. Definitions of concepts in epidemiological or clinical research correspond clinically to the categorization of individuals as having or being characterized by that concept. Successful definitions should give rise to *reliable* categorization of individuals, that is, identification of the

similar set of individuals, to enable stable (or consistent) descriptive and predictive estimates. Only then will inference, generalization, and clinical decision-making be possible. Using Monte Carlo methods to investigate psychometric properties of frailty indices, I showed that the number and composition of items in individual frailty indices strongly influence their reliability, even when using a single set of deficits. The point estimates and spreads for mean frailty scores, prevalence of frailty, cut-offs, mortality odds ratios, and predicted probability of mortality varied markedly between FIs with a low number of items and those with a high number of items. Frailty indices with 35 and 45 items were more reliable with intraclass coefficients of 0.75 and 0.84, respectively. However, as a group, FIs may not be sufficiently reliable to generalize results from one study or to compare one study to another. Frailty indices can characterize a portion of the increasing heterogeneity in aging, but the resulting clinical categorization may not be sufficiently reliable from a psychometric standpoint to allow direct clinical application.

In Manuscript 3, I explored a related but distinct theme: assuming that nosological constructs are sufficiently reliable for clinical implementation, how can the *clinical applicability* of a study be determined? Based on the existing literature, contemporary issues and debates, and brainstorming discussions, a framework was proposed to appraise the clinical applicability of studies. The framework relies on six criteria (*V*alidity, *I*ndication-informativeness, *C*linical relevance, *O*riginality, *R*isk-benefit comprehensiveness, and *T*ransposability) to classify studies and their findings into being non-practice informing (i.e., research findings), practice informing, or practice changing. In addition to being valid, a study should produce results that can inform an indication in order to *inform* practice. The four other criteria are further required to *change* practice. This proposed framework can be used formally assess whether studies featuring specific

constructs, namely frailty, are clinically applicable and to identify reasons for which they might not be.

In Manuscript 4, I investigated the clinical applicability of recent publications on frailty. This work touches upon issues related to the *domains of heterogeneity* captured in the multiple definitions of frailty, to the *measurement* of frailty as cut-offs used in defining frailty, and to direct *clinical applicability* as determined by mapping the function of frailty in 476 articles featuring frailty and appraising 150 articles in detail. Of 476 randomly selected articles published in 2017 and 2018, 150 (31%) used frailty in functions that could inform clinical indication. Among those, 38 definitions were used to measure frailty using various cut-offs. Among all articles, 63 (13%) articles were practice informing, 11 (2%) potentially practice changing, and 1 (0.2%) clearly practice changing. Lack of well-defined intervention and identifiable effect (96%) or originality (83%) were predominant barriers to clinical applicability. I found that only a minority of recent frailty research provides direct evidence of applicability to practice.

8.2 Strengths and limitations

The greatest overall strength of this work is its interdisciplinary focus. Collectively, the manuscripts explore fundamental questions on the basis of nosological constructs, their measurement, and their relevance to clinical practice.

Manuscript 1 is strengthened by the high quality and contemporary data from the CLSA. Methods were innovatively applied to quantify deviation (using absolute deviation from predicted mean models), assess the variation of deviation by age group, adjust for potential mean-variance relationships (when the mean was associated with chronological age), examine measurement properties (normative-clinical scaling), and investigate heterogeneity by domains (effective variance). I also formally examined the within-group variation of 85-year-olds to the between age-

group variation. Limitations include the exploratory nature of the analyses and potential issues of generalizability due to the exclusion of institutionalized and/or cognitively impaired older adults. Additionally, analyses were cross-sectional and could not disentangle period or cohort effects from the true aging process *per se*.[136]

The major strength of Manuscript 2 is the application of Monte Carlo methods to examine the reliability of measurements thus providing clinically oriented estimates of reliability (i.e., stability of descriptive and predictive estimates). I employed simulation methods in a single cohort of older adults with a unique set of 70 health deficit items to isolate the effect of varying the number and composition of items and maximize its potential reliability. As Manuscript 2 also used the CLSA participants (but restricting to those 65 years and older), the same limitation of generalizability applies due to the exclusion of institutionalized and/or cognitively impaired adults. Although having selected a unique set of 70 items and cut-offs to create frailty indices is a strength to ensure the highest potential reliability, it may also be a limitation as it may *overestimate* the actual reliability of frailty indices in usage which do not rely on similar items and cut-offs. Also, because the exact date of death was not available at the time of these analyses, predictive analyses between frailty and mortality using logistic regression models must be interpreted with caution, although the comparisons remain valid between estimates of each configuration.

The methods used to develop and refine the appraisal framework proposed in Manuscript 3 are the major strength of this manuscript. I examined the current literature and integrated contemporary debates and brainstorming discussions to develop a first version of the framework that was iteratively refined by testing. Although formal Delphi methodology was not employed, the process involved interdisciplinary contributions from clinicians, researchers, and methodologists with expertise in multiple substantive domains of clinical practice and research, as

well as epidemiology, biostatistics, qualitative, and translational research. Important limitations included, first, the lack of a formal systematic review process for appraising and synthesizing articles. Second, although we reached consensus on the proposed framework, there remains potential for subjectivity and variability in the qualitative process of reaching a consensus. Third, even if the importance of patient-centered outcomes is highlighted in the framework, our approach did not include direct patient involvement.

Two main strengths of Manuscript 4 are the use of systematic sampling to capture a representative sample of recent articles on frailty, and the use of our novel framework developed to specifically appraise the clinical applicability of frailty studies. Three limitations should be discussed. First, as for Manuscript 3, our mapping review may be subject to reviewer variability, even if we used pairs of independent reviewers for each article. Second, we sampled *published* articles from 2017 or 2018: inclusion of a wider chronological range of articles might have altered the results. Third, the mapping review focused only on individuals with frailty (i.e., those already considered frail), I did not fully appraise articles where frailty was used as an outcome or as a confounder. Fourth, I focused on the clinical use of frailty, not on its use in the research setting.

8.3 IMPLICATION OF FINDINGS FOR FRAILTY

Implications of findings and recommendations were presented in Table 4.3, Supplementary Table 4.5, Table 5.3, and Table 7.3. Below, I briefly summarize the implications for frailty and explore how they might extend to nosological constructs more broadly.

The findings presented have direct implications on frailty research and clinical usage, which may be distinguished by definitional and measurement implications, reporting and interpretation implications, and applicability implications. First, many frailty definitions currently coexist, and all definitions have a valid claim to capture relevant parts of the heterogeneity that

arises with chronological aging. However, the concurrent existence of definitions that encompass distinct domains of heterogeneity is problematic since it does not allow frailty to identify similar individuals as having frailty. What is frailty for one may not be frailty for another.[2–5] As I have argued, nosological constructs exist to inform and especially alter practice (either in the present or as a means for research and discovery for eventual application). This altering of clinical practice based on frailty is potentially specific to each definition. If frailty is about physical capacity versus multimorbidity (or distinct from it),[58] the relevant clinical action to respond to it will probably differ. Using frailty in practice, as any construct, requires the development of the best measure to identify it *according to what should follow its identification*. It cannot be the case that a single definition of frailty will encompass indications for all the potential interventions entertained for it. Specific interventions, and more largely specific frailty functions, will require specific "frailties." In the meantime, researchers and clinicians should carefully consider the specific definition and measurement of frailty used in each study.

Second, the current variability in definitions of frailty and their impact on reliability and stability of estimates has implications on the reporting and interpretation of frailty studies. The components and items of frailty measurements should be reported in detail to allow full replication of the measurement in other populations. Although many cut-offs for frailty measurements and for items composing them are often determined by a distributional cut-off (e.g., lowest quintile), the actual cut-off levels should be reported for application or replication in another population, to ensure reliable identification of those with that "instance" of frailty. Reporting and using actual values rather than distributional cut-offs anchors frailty measurements on a context-robust scale. Different definitions using different cut-offs also alter the consistency of associations between frailty "writ large" and outcomes. Although most studies on frailty reported positive associations

with outcomes (79% of studies in my scoping review of 150 articles), the combination of wide variability in definitions (including cut-offs), potential interventions per definitions, and outcomes challenges attempts to synthesize the body of evidence on frailty. Moreover, when frailty is mismeasured, bias may occur and alter the magnitude of associations with outcomes. Due to these many caveats, interpreting frailty as a unified construct may be impossible.

Third, although there are calls for implementing frailty into clinical practice, my findings highlight that, notwithstanding definitional and measurement issues, only a few studies on frailty have the potential to alter practice, i.e., to achieve the clinical imperative of improving health outcomes for older adults beyond current standard practice. Only a minority of appraised studies directly *indicated an intervention* to be coupled with the identification of "frailty." Of those that did, most were not *original* by lacking clinical or statistical significance or novelty. Ultimately, decisive success with regard to the clinical imperative for "frailty" would be to implement a *reliable definition* of frailty for which there is an *intervention*, and for which the *definition-intervention coupling* improves *patient-centered health outcomes*, *beyond* what would be achieved by *current standard practice*. To achieve this critical goal, matters of definitions, reporting, interpretation of frailty should be resolved before hastening its application to practice.

8.4 Implications and future directions

Implications discussed for frailty generalize beyond frailty and apply to other constructs. Fundamentally, nosological constructs characterize individuals in order to take a clinical action with the expectation of ameliorating health outcomes. The road between some constructs and health outcomes may be more tortuous than others. How the identification of HIV (both epistemically and clinically), for example, has led to improved health outcomes is clear; the case of frailty is less clear-cut. The work conducted in this dissertation sheds light on how to improve

the transition from coining a concept to fulfilling its clinical goal. First, what heterogeneity that exists should be characterized and how this heterogeneity arises in relation to the reference group selected, measurement (variable and domains measured) and scaling of measurement should be carefully explored. For example, abnormalities in vital signs,[137] laboratory values,[138,139] and bone mineral density are determined by distributional cut-offs, without having deliberately examined whether this provides the most optimal road to improving health outcomes in specific populations.

Second, many nosological constructs (in geriatrics, but also in other medical disciplines such as psychiatry[140]) are multidimensional and capture differing domains of heterogeneity. The questions raised by the numerous domains included in frailty also apply to the multifaceted nature of other age-related constructs such as polypharmacy, falls, or delirium. Classical medical syndromes have a single underlying cause but protean manifestations; geriatric syndromes do not require a *single* cause and rather focus on a single unifying *manifestation*. There is nothing inherently correct or incorrect about the dimensionality of a construct; what is important—from a clinical practice standpoint—is that the construct is defined in the most optimal way so to entail the intervention (understood very generally) that should follow. If polypharmacy calls for medication appraisal, is it best defined as taking *five or more* drugs?[14] If the presence of frailty suggests physical activity intervention, does frailty need to include affective, cognitive, and social support dimensions? Falls or delirium may entail specific interventions that arise directly in response: walking aids for falls and delirium-bundle interventions. But both also importantly require addressing the *underlying* causes which are interventions aimed at other elements causing or precipitating falls or delirium (e.g., cognitive impairment, acute medical conditions). From a clinical and functional perspective, many apparently unitary nosological constructs are actually decomposed into addressable components to be intervened on. This deconstruction or

decomposition between domain assessed and potential intervention may be a promising area of research arising from close attention to domains of heterogeneity and the concern with clinical applicability. Multidomain constructs may be most relevant when they are at the correct level of realization for potential interventions for it; interactions between or "mechanistic" co-occurrence of components may point to multidimensionality.

Nosological constructs are usually considered to be immutable in their definition. For example, depression is expected to be defined in the same manner whether it is used as a selection criterion, exposure, a mediator, a confounder, or an outcome. However, there may be more relevant constructs (and measures) of depression if it is selection criterion for a drug trial, as an exposure for mortality, or an outcome for a trial of physical activity.[118,140,141] There is some leeway in modeling exposure, confounders (e.g., propensity scores which use multiple variables), [107–109] and outcome variables in epidemiological studies. A third ontological implication is that the development and analysis of constructs may benefit from being more informed by the expected nosological function. In clinical epidemiology, constructs and concepts are foremost tools to influence clinical practice. Though reliability in use is essential, reification is not: definitions in the absolute may be less powerful than functional definitions tailored to specific contexts. Much of the current methodological research conducted in clinical epidemiology concerns issues of measurements, psychometrics, diagnostics, advanced modeling, evidence review and synthesis (meta-research); perhaps ontological issues, such as *what* is measured, *why* it is measured, and *how* to best develop useful nosological constructs may be a promising area of research. Is the ideal of clinical epidemiology to precisely quantify relationships between health variables or rather to precisely determine what is worth finding relationships about? For frailty and other constructs,

one may wonder, having picked most low-hanging fruits from current methods, whether it would be time to tend to the tree we are picking from—or better yet, find another tree in the forest.

8.5 Conclusion

In the last fifty years, epidemiological research on aging and geriatric medicine have gradually set their sights on frailty and other age-related nosological constructs. Keeping in mind the clinical imperative to improve health of populations and individuals, this dissertation explored the basis (i.e., heterogeneity) of age-related constructs, the stability (i.e., reliability) of frailty categorization based on the frailty deficit-accumulation index, the criteria by which categorizations can inform or alter clinical practice (i.e., clinical applicability), and the recent applicability of frailty to clinical practice. Heterogeneity increases with aging for some health characteristics allowing for potential differential categorization of older adults. However, categorizations may be hampered by low reliability and thus generalizability. Not all categorizations and studies can equally inform or alter clinical practice. Frailty, as an example of one such categorization, may not be directly clinically applicable due to definition, indication-informativeness, and originality challenges. Research on aging and geriatric medicine are young disciplines. By providing an overarching framework to analyze age-related constructs, my hope is to have contributed to set ideas at the intersection of research and applicability to clinical practice. I hope that the work here presented will not lead to a frailty cul-de-sac, but rather to renewed avenues of research and practice implementation less marred by imprecision or confusion. Whether I have erred or not, I remain firmly convinced that "truth emerges more readily from error than from confusion."[128]

References

1. Clegg A, Young J, Iliffe S, Rikkert MO, Rockwood K. Frailty in elderly people. *Lancet*. 2013;381(9868):752-762. doi:10.1016/S0140-6736(12)62167-9

2. Dent E, Kowal P, Hoogendijk EO. Frailty measurement in research and clinical practice: a review. *Eur J Intern Med*. 2016;31:3-10. doi:10.1016/j.ejim.2016.03.007

3. Rodríguez-Mañas L, Féart C, Mann G, et al. Searching for an operational definition of frailty: A delphi method based consensus statement. the frailty operative definition-consensus conference project. *J Gerontol A Biol Sci Med Sci*. 2013;68(1):62-67. doi:10.1093/gerona/gls119

4. Aguayo GA, Donneau AF, Vaillant MT, et al. Agreement between 35 published frailty scores in the general population. *Am J Epidemiol*. 2017;186(4):420-434. doi:10.1093/aje/kwx061

5. Xue Q-L, Bandeen-Roche K, Varadhan R, Zhou J, Fried LP. Initial Manifestations of Frailty Criteria and the Development of Frailty Phenotype in the Women's Health and Aging Study. *Journals Gerontol Ser A Biol Sci Med Sci*. 2008;63(9):984-990. doi:10.1093/gerona/63.9.984

6. Bandeen-Roche K, Seplaki CL, Huang J, et al. Frailty in older adults: A nationally representative profile in the United States. *J Gerontol A Biol Sci Med Sci*. 2015;70(11):1427-1434. doi:10.1093/gerona/glv133

7. Buta BJ, Walston JD, Godino JG, et al. Frailty assessment instruments: systematic characterization of the uses and contexts of highly-cited instruments. *Ageing Res Rev*. 2016;26:53-61. doi:10.1016/j.arr.2015.12.003

8. Theou O, Squires E, Mallery K, et al. What do we know about frailty in the acute care setting? A scoping review. *BMC Geriatr*. 2018;18(1). doi:10.1186/s12877-018-0823-2

9. Morley JE, Vellas B, van Kan GA, et al. Frailty Consensus: A Call to Action. *J Am Med Dir Assoc*. 2013;14(6):392-397. doi:10.1016/j.jamda.2013.03.022

10. Cesari M, Prince M, Thiyagarajan JA, et al. Frailty: An Emerging Public Health Priority. *J*

Am Med Dir Assoc. 2016;17(3):188-192. doi:10.1016/j.jamda.2015.12.016

11. Mudge AM, Hubbard RE. Frailty: Mind the gap. *Age Ageing*. 2018;47(4):508-511. doi:10.1093/ageing/afx193

12. Fabbri E, Zoli M, Gonzalez-Freire M, Salive ME, Studenski SA, Ferrucci L. Aging and multimorbidity: new tasks, priorities, and frontiers for integrated gerontological and clinical research. *J Am Med Dir Assoc*. 2015;16(8):640-647. doi:10.1016/j.jamda.2015.03.013

13. Nguyen QD, Wu C, Odden MC, Kim DH. Multimorbidity patterns, frailty, and survival in community-dwelling older adults. *J Gerontol A Biol Sci Med Sci*. 2019;74(8):1265-1270. doi:10.1093/gerona/gly205/5088142

14. Fincke BG, Snyder K, Cantillon C, et al. Three complementary definitions of polypharmacy: Methods, application and comparison of findings in a large prescription database. *Pharmacoepidemiol Drug Saf*. 2005;14(2):121-128. doi:10.1002/pds.966

15. Oh ES, Fong TG, Hshieh TT, Inouye SK. Delirium in older persons: Advances in diagnosis and treatment. *JAMA - J Am Med Assoc*. 2017;318(12):1161-1174. doi:10.1001/jama.2017.12067

16. Quinlan N, Marcantonio ER, Inouye SK, Gill TM, Kamholz B, Rudolph JL. Vulnerability: The crossroads of frailty and delirium. *J Am Geriatr Soc*. 2011;59(SUPPL. 2):262-268. doi:10.1111/j.1532-5415.2011.03674.x

17. Inouye SK, Studenski S, Tinetti ME, Kuchel GA. Geriatric syndromes: clinical, research, and policy implications of a core geriatric concept. *J Am Geriatr Soc*. 2007;55(5):780-791. doi:10.1111/j.1532-5415.2007.01156.x

18. Morley JE. The New Geriatric Giants. *Clin Geriatr Med*. 2017;33(3):xi-xii. doi:10.1016/j.cger.2017.05.001

19. Tinetti ME, Fried T. The end of the disease era. *Am J Med*. 2004;116(3):179-185. doi:10.1016/j.amjmed.2003.09.031

20. Diez Roux A V. The Unique Space of Epidemiology: Drawing on the Past to Project into

the Future. *Am J Epidemiol*. 2019;188(5):886-889. doi:10.1093/aje/kwz001

21. Fried LP, Walston JD, Ferrucci L. Frailty. In: Halter JB, Ouslander JG, Tinetti ME, Studenski S, High KP, Asthana S, eds. *Hazzard's Geriatric Medicine and Gerontology*. 6th ed. McGraw-Hill; 2009:631-645.

22. Sternberg SA, Wershof Schwartz A, Karunananthan S, Bergman H, Mark Clarfield A. The identification of frailty: a systematic literature review. *J Am Geriatr Soc*. 2011;59(11):2129-2138. doi:10.1111/j.1532-5415.2011.03597.x

23. Gobbens RJ, Luijkx KG, Wijnen-Sponselee MT, Schols JM. Toward a conceptual definition of frail community dwelling older people. *Nurs Outlook*. 2010;58(2):76-86. doi:10.1016/j.outlook.2009.09.005

24. Ferrucci L, Fabbri E, Walston JD. Frailty. In: Halter JB, Ouslander JG, Studenski S, et al., eds. *Hazzard's Geriatric Medicine and Gerontology*. 7th ed. McGraw-Hill; 2017:691-708.

25. Rockwood K. What would make a definition of frailty successful? *Age Ageing*. 2005;34(5):432-434. doi:10.1093/ageing/afi146

26. Hoogendijk EO, Afilalo J, Ensrud KE, Kowal P, Onder G, Fried LP. Series Frailty 1 Frailty : implications for clinical practice and public health. *Lancet*. 2019;394(10206):1365-1375. doi:10.1016/S0140-6736(19)31786-6

27. Cesari M, Pérez-Zepeda MU, Marzetti E. Frailty and Multimorbidity: Different Ways of Thinking About Geriatrics. *J Am Med Dir Assoc*. 2017;18(4):361-364. doi:10.1016/j.jamda.2016.12.086

28. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology*. 1999;10(1):37-48. doi:10.1097/00001648-199901000-00008

29. Pearl J. *Causality*. 2nd Editio. New York, NY: Cambridge University Press; 2009.

30. Holland PW. Statistics and Casual Inference. *J Am Stat Assoc*. 1986;81(396):945-960. doi:10.1080/01621459.1986.10478354

31. Xue Q-L, Varadhan R. What Is Missing in the Validation of Frailty Instruments?..."Frailty Consensus: A Call to Action" by Morley et al. *J Am Med Dir Assoc*. 2014;15(2):141-143.

doi:10.1016/j.jamda.2013.11.007

32. Old and frail. *Br Med J*. 1968;1(5594):723 LP - 724. doi:10.1136/bmj.1.5594.723

33. Vaupel JW, Manton KG, Stallard E. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*. 1979;16(3):439. doi:10.2307/2061224

34. Rockwood K, Fox R, Stolee P. Frailty in elderly people: an evolving concept. *C Can*. 1994;150(4):489-495. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1486322/. Accessed June 18, 2016.

35. Stallones RA. To advance epidemiology. *Annu Rev Public Health*. 1980;1:69-82. doi:10.1146/annurev.pu.01.050180.000441

36. Fried LP, Borhani NO, Enright P, et al. The cardiovascular health study: Design and rationale. *Ann Epidemiol*. 1991;1(3):263-276. doi:10.1016/1047-2797(91)90005-W

37. The Canadian Study of Health and Aging. The Canadian Study of Health and Aging, Risk factors for Alzheimer's disease in Canada. *Neurology*. 1994;150(6):899-913.

38. Walston JD, Bandeen-Roche K. Frailty: a tale of two concepts. *BMC Med*. 2015;13:185. doi:10.1186/s12916-015-0420-6

39. Fried LP, Tangen CM, Walston J, et al. Frailty in older adults: evidence for a phenotype. *J Gerontol A Biol Sci Med Sci*. 2001;56(3):808-813. doi:10.1093/gerona/56.3.M146

40. Mitnitski AB, Mogilner AJ, Rockwood K. Accumulation of deficits as a proxy measure of aging. *Sci World*. 2001;1:323-336. doi:10.1100/tsw.2001.58

41. Rockwood K, Song X, MacKnight C. A global clinical measure of fitness and frailty. *Can Med Assoc J*. 2005;173(5):489-495. https://www.cmaj.ca/content/173/5/489.full. Accessed July 13, 2016.

42. Searle SD, Mitnitski A, Gahbauer EA, Gill TM, Rockwood K. A standard procedure for creating a frailty index. *BMC Geriatr*. 2008;8(1):24. doi:10.1186/1471-2318-8-24

43. de Vries NM, Staal JB, van Ravensberg CD, Hobbelen JSM, Rikkert MGMO, der Sanden MWGN. Outcome instruments to measure frailty: a systematic review. *Ageing Res Rev*. 2011;10(1):104-114. doi:10.1016/j.arr.2010.09.001

44. Coltman T, Devinney TM, Midgley DF, Venaik S. Formative versus reflective measurement models: Two applications of formative measurement. *J Bus Res*. 2008;61(12):1250-1262. doi:https://doi.org/10.1016/j.jbusres.2008.01.013

45. Kuchel GA. Frailty and Resilience as Outcome Measures in Clinical Trials and Geriatric Care: Are We Getting Any Closer? *J Am Geriatr Soc*. 2018. doi:10.1111/jgs.15441

46. Afilalo J, Alexander KP, Mack MJ, et al. Frailty assessment in the cardiovascular care of older adults. *J Am Coll Cardiol*. 2014;63(8):747-762. doi:10.1016/j.jacc.2013.09.070

47. Guralnik JM, Simonsick EM, Ferrucci L, et al. A Short Physical Performance Battery Assessing Lower Extremity Function: Association With Self-Reported Disability and Prediction of Mortality and Nursing Home Admission. *J Gerontol*. 1994;49(2):M85-M94. doi:10.1093/geronj/49.2.M85

48. Rolfson DB, Majumdar SR, Tsuyuki RT, Tahir A, Rockwood K. Validity and reliability of the Edmonton Frail Scale. *Age Ageing*. 2006;35(5):526-529. doi:10.1093/ageing/afl041

49. Gobbens RJJ, van Assen MALM, Luijkx KG, Wijnen-Sponselee MT, Schols JMGA. The Tilburg frailty indicator: Psychometric properties. *J Am Med Dir Assoc*. 2010;11(5):344-355. doi:10.1016/j.jamda.2009.11.003

50. Gilbert T, Neuburger J, Kraindler J, et al. Development and validation of a Hospital Frailty Risk Score focusing on older people in acute care settings using electronic hospital records: an observational study. *Lancet*. 2018;391(10132):1775-1782. doi:10.1016/S0140-6736(18)30668-8

51. McIsaac DI, Bryson GL, van Walraven C. Association of Frailty and 1-Year Postoperative Mortality Following Major Elective Noncardiac Surgery. *JAMA Surg*. 2016;151(6):538. doi:10.1001/jamasurg.2015.5085

52. Cesari M, Theou O. Frailty: The Broad View. In: Fillit HM, Rockwood K, Young JB, eds. *Brocklehurst's Textbook of Geriatric Medicine and Gerontology*. 8th ed. Elsevier Health Sciences; 2016:82-87.

53. Fodor JA. Special sciences (or: The disunity of science as a working hypothesis). *Synthese*. 1974;28(2):97-115. doi:10.1007/BF00485230

54. Wang J, Maxwell CA, Yu F. Biological processes and biomarkers related to frailty in older adults: a state-of-the-science literature review. *Biol Res Nurs*. 2019;21(1):80-106. doi:10.1177/1099800418798047

55. Handforth C, Clegg A, Young C, et al. The prevalence and outcomes of frailty in older cancer patients: a systematic review. *Ann Oncol*. 2015;26(6):1091-1101. doi:10.1093/annonc/mdu540

56. Wu C, Smit E, Xue Q-L, Odden MC. Prevalence and Correlates of Frailty Among Community-Dwelling Chinese Older Adults: The China Health and Retirement Longitudinal Study. *J Gerontol A Biol Sci Med Sci*. 2017;73(1):102-108. doi:10.1093/gerona/glx098

57. Ferrucci L, Windham BG, Fried LP. Frailty in older persons. *Genus*. 2005;61(1):39-53. http://www.jstor.org/stable/29788835. Accessed June 17, 2017.

58. Fried LP, Ferrucci L, Darer J, Williamson JD, Anderson G. Untangling the concepts of disability, frailty, and comorbidity: implications for improved targeting and care. *J Gerontol A Biol Sci Med Sci*. 2004;59(3):255-263. doi:10.1093/gerona/59.3.M255

59. Berkman LF, Kawachi I, Glymour MM. *Social Epidemiology*. Oxford, UK: Oxford University Press; 2015. doi:10.1093/med/9780195377903.001.0001

60. Diez-Roux A V. Multilevel Analysis in Public Health Research. *Annu Rev Public Health*. 2000;21(1):171-192. doi:10.1146/annurev.publhealth.21.1.171

61. Xue Q-L. The Frailty Syndrome: Definition and Natural History. *Clin Geriatr Med*. 2011;27(1):1-15. doi:10.1016/j.cger.2010.08.009

62. Radloff LS. The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. *Appl Psychol Meas*. 1977;1(3):385-401. doi:10.1177/014662167700100306

63. Abellan van Kan G, Rolland YM, Morley JE, Vellas B. Frailty: Toward a Clinical Definition. *J Am Med Dir Assoc*. 2008;9(2):71-72. doi:10.1016/j.jamda.2007.11.005

64. Morley J, Malmstrom T, Miller D. A simple frailty questionnaire (FRAIL) predicts

outcomes in middle aged African Americans. *J Nutr Heal*. 2012. http://link.springer.com/article/10.1007/s12603-012-0084-2. Accessed June 21, 2016.

65. Ensrud KE. Comparison of 2 Frailty Indexes for Prediction of Falls, Disability, Fractures, and Death in Older Women. *Arch Intern Med*. 2008;168(4):382. doi:10.1001/archinternmed.2007.113

66. Ensrud KE, Ewing SK, Cawthon PM, et al. A comparison of frailty indexes for the prediction of falls, disability, fractures, and mortality in older men. *J Am Geriatr Soc*. 2009;57(3):492-498. doi:10.1111/j.1532-5415.2009.02137.x

67. Rockwood K, Mitnitski A. Aging and deficit accumulation: clinical implications. In: Fillit HM, Rockwood K, Young J, eds. *Brocklehurst's Textbook of Geriatric Medicine and Gerontology*. 8th Editio. ; 2017.

68. Jones DM, Song X, Rockwood K. Operationalizing a frailty index from a standardized comprehensive geriatric assessment. *J Am Geriatr Soc*. 2004;52(11):1929-1933. doi:10.1111/j.1532-5415.2004.52521.x

69. Clegg A, Bates C, Young J, et al. Development and validation of an electronic frailty index using routine primary care electronic health record data. *Age Ageing*. 2016;45(3):353-360. doi:10.1093/ageing/afw039

70. Howlett SE, Rockwood MRH, Mitnitski A, Rockwood K. Standard laboratory tests to identify older adults at increased risk of death. *BMC Med*. 2014;12:171. doi:10.1186/s12916-014-0171-9

71. Pajewski NM, Williamson JD, Applegate WB, et al. Characterizing Frailty Status in the Systolic Blood Pressure Intervention Trial. *J Gerontol A Biol Sci Med Sci*. 2016;71(5):649-655. doi:10.1093/gerona/glv228

72. Sanders NA, Supiano MA, Lewis EF, et al. The frailty syndrome and outcomes in the TOPCAT trial. *Eur J Heart Fail*. 2018;20(11):1570-1577. doi:10.1002/ejhf.1308

73. Subramaniam S, Aalberg JJ, Soriano RP, Divino CM. New 5-Factor Modified Frailty Index Using American College of Surgeons NSQIP Data. *J Am Coll Surg*. 2018;226(2):173-181.e8. doi:10.1016/j.jamcollsurg.2017.11.005

74. Farhat JS, Velanovich V, Falvo AJ, et al. Are the frail destined to fail? Frailty index as predictor of surgical morbidity and mortality in the elderly. *J Trauma Acute Care Surg*. 2012;72(6):1526-1530; discussion 1530-1. doi:10.1097/TA.0b013e3182542fab

75. Drubbel I, Numans ME, Kranenburg G, Bleijenberg N, De Wit NJ, Schuurmans MJ. Screening for frailty in primary care: a systematic review of the psychometric properties of the frailty index in community-dwelling older people. *BMC Geriatr*. 2014;14(1):1-13. doi:10.1186/1471-2318-14-27

76. Romero-Ortuno R. An alternative method for Frailty Index cut-off points to define frailty categories. *Eur Geriatr Med*. 2013;4(5):299-303. doi:10.1016/j.eurger.2013.06.005

77. Pulok MH, Theou O, Valk AM Van Der, Rockwood K. The role of illness acuity on the association between frailty and mortality in emergency department patients referred to internal medicine. 2020;(May):1071-1079. doi:10.1093/ageing/afaa089

78. Raîche M, Hébert R, Dubois MF. PRISMA-7: A case-finding tool to identify older adults with moderate to severe disabilities. *Arch Gerontol Geriatr*. 2008;47(1):9-18. doi:10.1016/j.archger.2007.06.004

79. Muscedere J, Andrew MK, Bagshaw SM, et al. Screening for frailty in Canada's health care system: a time for action. *Can J Aging*. 2016;35(3):281-297. doi:10.1017/S0714980816000301

80. Diabetes Canada Clinical Practice Guidelines Expert Committee. Diabetes Canada 2018 Clinical Practice Guidelines for the Prevention and Management of Diabetes in Canada. *Can J Diabetes*. 2018;42(Suppl 1:S1-S325. doi:10.1080/00945718308059306

81. Dent E, Lien C, Lim WS, et al. The Asia-Pacific clinical practice guidelines for the management of frailty. *J Am Med Dir Assoc*. 2017;18(7):564-575. doi:10.1016/j.jamda.2017.04.018

82. Leung AA, Daskalopoulou SS, Dasgupta K, et al. Hypertension Canada's 2017 Guidelines for Diagnosis, Risk Assessment, Prevention, and Treatment of Hypertension in Adults. *Can J Cardiol*. 2017;33(5):557-576. doi:10.1016/j.cjca.2017.03.005

83. Laube R, Wang H, Park L, et al. Frailty in advanced liver disease. *Liver Int*.

2018;38(12):2117-2128.

84. Brothers TD, Kirkland S, Guaraldi G, et al. Frailty in People Aging With Human Immunodeficiency Virus (HIV) Infection. *J Infect Dis*. 2014;210(8):1170-1179. doi:10.1093/infdis/jiu258

85. McAdams-Demarco MA, Law A, Salter ML, et al. Frailty as a novel predictor of mortality and hospitalization in individuals of all ages undergoing hemodialysis. *J Am Geriatr Soc*. 2013;61(6):896-901. doi:10.1111/jgs.12266

86. Fragoso CAV, Enright PL, McAvay G, Van Ness PH, Gill TM. Frailty and respiratory impairment in older persons. *Am J Med*. 2012;125(1):79-86.

87. Walston J, Robinson TN, Zieman S, et al. Integrating frailty research into the medical specialties—report from a U13 Conference. *J Am Geriatr Soc*. 2017;65(10):2134-2139. doi:10.1111/jgs.14902

88. Makary MA, Segev DL, Pronovost PJ, et al. Frailty as a Predictor of Surgical Outcomes in Older Patients. *J Am Coll Surg*. 2010;210(6):901-908. doi:10.1016/j.jamcollsurg.2010.01.028

89. Statistics Canada. Seniors and aging statistics. https://www.statcan.gc.ca/eng/subjects-start/seniors_and_aging. Published 2021.

90. Nascher IL. Geriatrics. *N Y Med J*. 1909;(90):358-359.

91. Nascher IL. *Geriatrics: The Diseases of Old Age and Their Treatment*. Philadelphia: P. Blakistons's Son & Co; 1914.

92. Aronson L. Elderhood : redefining aging, transforming medicine, reimagining life. 2019. https://www.overdrive.com/search?q=235DF0F7-C22D-4870-A7BE-561E563CC400.

93. Parker SG, McCue P, Phelps K, et al. What is Comprehensive Geriatric Assessment (CGA)? An umbrella review. *Age Ageing*. 2018;47(1):149-155. doi:10.1093/ageing/afx166

94. Rothman KJ. Causes. *Am J Epidemiol*. 1976;104(6):587-592. http://aje.oxfordjournals.org/content/104/6/587.extract. Accessed January 18, 2016.

95. Rothman KJ, Greenland S. Causation and Causal Inference in Epidemiology. *Am J Public Health*. 2005;95(S1):S144-S150. doi:10.2105/AJPH.2004.059204

96. Cassell EJ. *The Nature of Suffering and the Goals of Medicine*. New York: Oxford University Press; 2004.

97. Isaacs B. *An Introduction to Geriatrics*. Williams & Wilkins Company; 1965.

98. Halter JB, Ouslander JG, Studenski S, et al. *Hazzard's Geriatric Medicine and Gerontology*.; 2017.

99. Fillit H, Rockwood K, Young J. Brocklehurst's textbook of geriatric medicine and gerontology. 2017. https://www.clinicalkey.com/#!/browse/book/3-s2.0-C20120076016.

100. Ahmed N, Mandel R, Fain MJ. Frailty: An Emerging Geriatric Syndrome. *Am J Med*. 2007;120(9):748-753. doi:10.1016/j.amjmed.2006.10.018

101. Flacker JM. What Is A Geriatric Syndrome Anyway? *J Am Geriatr Soc*. 2003;51(4):574-576. doi:10.1046/j.1532-5415.2003.51174.x

102. Tinetti ME, Inouye SK, Gill TM, Doucette JT. Shared Risk Factors for Falls, Incontinence, and Functional Dependence: Unifying the Approach to Geriatric Syndromes. *JAMA*. 1995;273(17):1348-1353. doi:10.1001/jama.1995.03520410042024

103. Olde Rikkert MGM, Rigaud AS, van Hoeyweghen RJ, de Graaf J. Geriatric syndromes: Medical misnomer or progress in geriatrics? *Neth J Med*. 2003;61(3):83-87.

104. Miller WT. Letter from the editor: Occam versus Hickam. *Semin Roentgenol*. 1998;33(3):213. doi:10.1016/S0037-198X(98)80001-1

105. Armstrong EC. The well-built clinical question: the key to finding the best evidence efficiently. *WMJ*. 1999;98(2):25-28.

106. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. 3rd Editio. Lippincott Williams & Wilkins; 2008.

107. Greenland S, Morgenstern H. Confounding in health research. *Annu Rev Public Heal*. 2001;22:189-212.

108. Greenland S, Pearce N. Statistical foundations for model-based adjustments. *Annu Rev Public Health*. 2015;36:89-108. doi:10.1146/annurev-publhealth-031914-122559

109. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70(1):41-55.

110. Galea S. An argument for a consequentialist epidemiology. *Am J Epidemiol*. 2013;178(8):1185-1191. doi:10.1093/aje/kwt172

111. Chiolero A. Counterfactual and interventionist approach to cure risk factor epidemiology. *Int J Epidemiol*. 2016;45(6):2202-2203. doi:10.1093/ije/dyw159

112. Verghese J, Annweiler C, Ayers E, et al. Motoric cognitive risk syndrome Multicountry prevalence and dementia risk. *Neurology*. 2014;83(8):718-726. doi:10.1212/WNL.0000000000000717

113. Van Den Akker M, Buntinx F, Knottnerus JA. Comorbidity or multimorbidity: What's in a name? A review of literature. *Eur J Gen Pract*. 1996;2(2):65-70. doi:10.3109/13814789609162146

114. Levine ME. Modeling the rate of senescence: Can estimated biological age predict mortality more accurately than chronological age? *Journals Gerontol - Ser A Biol Sci Med Sci*. 2013;68(6):667-674. doi:10.1093/gerona/gls233

115. Bynum WF. History of medicine : a very short introduction. 2008. http://site.ebrary.com/id/10464218.

116. Illich I. Medical nemesis : the expropriation of health. 1976. http://books.google.com/books?id=dnBHAAAAMAAJ.

117. Tikkinen KAO, Leinonen JS, Guyatt GH, Ebrahim S, Järvinen TLN. What is a disease? Perspectives of the public, health professionals and legislators. *BMJ Open*. 2012;2(6):e001632. doi:10.1136/bmjopen-2012-001632

118. Kendler KS. An historical framework for psychiatric nosology. *Psychol Med*. 2009;39(12):1935. doi:10.1017/S0033291709005753

119. Foucault M. *L'archéologie du savoir*. [Paris]: Gallimard; 1969.

120. Scully JL. What is a disease? *EMBO Rep*. 2004;5(7):650-653. doi:10.1038/sj.embor.7400195

121. Elstein AS. Thinking about diagnostic thinking: A 30-year perspective. *Adv Heal Sci Educ*. 2009;14(1 SUPPL):7-18. doi:10.1007/s10459-009-9184-0

122. Charlin B, Lubarsky S, Millette B, et al. Clinical reasoning processes: Unravelling complexity through graphical representation. *Med Educ*. 2012;46(5):454-463. doi:10.1111/j.1365-2923.2012.04242.x

123. Charlin B, Tardif J, Boshuizen HPA. Scripts and Medical Diagnostic Knowledge: Theory and Applications for Clinical Reasoning Instruction and Research. *Acad Med*. 2000;75(2). https://journals.lww.com/academicmedicine/Fulltext/2000/02000/Scripts_and_Medical_Diagnostic_Knowledge__Theory.20.aspx.

124. Cook DA, Sherbino J, Durning SJ. Management reasoning beyond the diagnosis. *JAMA - J Am Med Assoc*. 2018;319(22):2267-2268. doi:10.1001/jama.2018.4385

125. Theou O, Walston J, Rockwood R. Operationalizing Frailty Using the Frailty Phenotype and Deficit Accumulation Approaches. *Frailty Aging Biol Clin Soc Implic*. 2015;108:66-73. doi:10.1159/000381164

126. Rockwood K, Mitnitski A. Unconventional Views of Frailty Frailty in Relation to the Accumulation of Deficits. *J Gerontol A Biol Sci Med Sci*. 2007;62(7):722-727.

127. Wittgenstein L. *Philosophical investigations = philosophische untersuchungen*. Oxford: Basil Blackwell; 1953.

128. Bacon F, Kitchin GW. *The Novum Organon; or, A True Guide to the Interpretation of Nature*. Oxford: University Press; 1855.

129. Wikipedia - Wikimedia Foundation. Medicine. https://en.wikipedia.org/w/index.php?title=Medicine&oldid=1022412735. Published 2021.

130. Raina PS, Wolfson C, Kirkland SA, et al. The Canadian Longitudinal Study on Aging (CLSA). *Can J Aging*. 2009;28(3):221-229. doi:10.1017/s0714980809990055

131. Raina P, Wolfson C, Kirkland S, et al. Cohort Profile: The Canadian Longitudinal Study on Aging (CLSA). *Int J Epidemiol*. 2019;48(6):1752-1753j. doi:10.1093/ije/dyz173

132. Peña D, Rodríguez J. Descriptive measures of multivariate scatter and linear dependence. *J Multivar Anal*. 2003;85(2):361-374. doi:10.1016/S0047-259X(02)00061-1

133. Peña D, Van Der Linde A. Dimensionless measures of variability and dependence for multivariate continuous distributions. *Commun Stat - Theory Methods*. 2007;36(10):1845-1854. doi:10.1080/03610920601126449

134. Streiner DL, Norman GR, Cairney J. *Health Measurement Scales A Practical Guide to Their Development and Use*. 5th ed. Oxford, UK: Oxford University Press; 2014. doi:10.1093/med/9780199685219.001.0001

135. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. *Measurement in Medicine: A Practical Guide*. Cambridge: Cambridge University Press; 2011. doi:DOI: 10.1017/CBO9780511996214

136. Dannefer D, Sell RR. Age structure, the life course and "aged heterogeneity": prospects for research and theory. *Compr Gerontol B*. 1988;2(1):1-10. http://www.ncbi.nlm.nih.gov/pubmed/3180151.

137. Mackowiak PA. A critical appraisal of 98.6 degrees F, the upper limit of the normal body temperature, and other legacies of Carl Reinhold August Wunderlich. *JAMA J Am Med Assoc*. 1992;268(12):1578-1580. doi:10.1001/jama.268.12.1578

138. Jones GRD, Haeckel R, Loh TP, et al. Indirect methods for reference interval determination - Review and recommendations. *Clin Chem Lab Med*. 2019;57(1):20-29. doi:10.1515/cclm-2018-0073

139. Farrell CJL, Nguyen L. Indirect reference intervals: Harnessing the power of stored laboratory data. *Clin Biochem Rev*. 2019;40(2):99-111. doi:10.33176/AACB-19-00022

140. Kendler KS, Charles O. Gardner J. Boundaries of Major Depression: An Evaluation of DSM-IV Criteria. *Am J Psychiatry*. February 1998. http://ajp.psychiatryonline.org/doi/abs/10.1176/ajp.155.2.172. Accessed December 13, 2015.

141. Kendell R, Jablensky A. Distinguishing Between the Validity and Utility of Psychiatric Diagnoses. *Am J Psychiatry*. 2003;160(1):4-12. doi:10.1176/appi.ajp.160.1.4