Assessment of cognition in ageing. Investigating internal validity, occurrence and reversion of Mild Cognitive Impairment. Data from the general population study “Good Aging in Skåne”

Overton, Marie Claire

2019

Document Version:
Publisher's PDF, also known as Version of record

Assessment of cognition in ageing. Investigating internal validity, occurrence and reversion of Mild Cognitive Impairment

Data from the general population study "Good Aging in Skåne"

MARIECLAIRE OVERTON
GERIATRICS | FACULTY OF MEDICINE | LUND UNIVERSITY

MARIECLAIRE OVERTON has previously worked as a test administrator in the Good Aging in Skåne study. She has a master's degree in psychology. Her thesis explores methodological aspects of research on cognition and ageing. Focus lies on test administrator influence, birth cohort effects and the occurrence, reversion and operationalisation of Mild Cognitive Impairment.

Printed by Media-Tryck, Lund 2019 NORDIC SWAN ECOLABEL 3041 0903

LUND UNIVERSITY

FACULTY OF MEDICINE

Department of Clinical Sciences
Geriatrics

Lund University, Faculty of Medicine
Doctoral Dissertation Series 2019:69
ISBN 978-91-7619-798-1
ISSN 1652-8220

Assessment of cognition in ageing. Investigating internal validity, occurrence and reversion of Mild Cognitive Impairment

Assessment of cognition in ageing. Investigating internal validity, occurrence and reversion of Mild Cognitive Impairment

Data from the general population study "Good Aging in Skåne"

Marieclaire Overton

LUND
UNIVERSITY

DOCTORAL DISSERTATION

by due permission of the Faculty of Medicine, Lund University, Sweden.
To be defended at the Clinical Research Centre Aula, Jan Waldenströms gata 35, Skåne University hospital Malmö. Friday 6th of September 2019

Faculty opponent
Professor Ingmar Skoog, Gothenburg University, Sweden

Organization LUND UNIVERSITY	Document name DOCTORAL DISSERTATION
	Date of issue 6th of September 2019
Author: Marieclaire Overton	Sponsoring organization

Title and subtitle: Assessment of cognition in ageing. Investigating internal validity, occurrence and reversion of Mild Cognitive Impairment. Data from the general population study "Good Aging in Skåne".

Abstract

Given a growing population of older adults and the common occurrence of natural as well as pathological cognitive decline, i.e. dementia, there is a need to better understand how to accurately measure and classify level of cognitive functioning. Both using multiple test administrators to collect cognitive data and what year the participant is born (birth cohort effects) can influence participants' performance on cognitive tests. The concept Mild cognitive impairment (MCI) aims to capture a prodromal stage of dementia. There is heterogeneity in the reported MCI prevalence and incidence estimates, partly due to varying MCI definitions applied in the research literature. Also, estimations of reversion rates (recovering from MCI) are reported to be fairly high (≈29%). Sampled participants used in these studies, aged 60-100, were randomly invited to partake in the Swedish population-based ageing study: *Good Aging in Skåne*.

Paper I (*n* of participants =6686, M_{age}=71.3, *n* of test administrators = 21) examined test administrator influence on participants' test scores. A series of mixed linear models revealed significant random effects corresponding to the test administrators for cognitive test measuring episodic memory, speed of processing and spatial ability (p<.01). The variation seen in test scores ascribed to the test administrator was between 1.4%-3.5%.

Paper II (*n* of 60-year olds = 736 born between 1942-55, *n* of 81-year olds = 431 born between 1920-33) examined differences in participants' tests scores in 3 sets of cohorts, at age 60 or 81, respectively. The ANOVAs found significant (p<0.05) differences in test scores measuring speed of processing, episodic memory, attention, executive functioning and vocabulary between the birth cohorts, where the later born cohorts outperformed the earlier born cohorts.

Paper III (*n* for prevalence =3752, *n* for incidence = 1451, age-groups= 60-69, 70-79, 80+) examined prevalence and incidence of MCI across age, sex, subtypes and criteria of cognitive impairment. Prevalence of MCI were 21.4% and 6.6% for a lenient and strict inclusion criterion of cognitive impairment, respectively. The MCI incidence rates were 22.6 and 8.67 per 1000 person-years for a lenient and strict criterion, respectively. No sex-differences in MCI estimates were established and age differences in estimates were inconclusive.

Paper IV (*n*=331) examined a 6-year MCI reversion rate and investigated factors that predicted reversion. There was a high reversion rate of 58%. The logistic regression found that lower age (p<0.05), better global cognitive functioning (p<0.02), good concentration (p<0.05) and single domain subtype (p<0.001) could predict reversion.

In summary, test administrator effects and birth cohort effects were present in the cognitive data of this population based ageing study. Moreover, prevalent MCI was fairly common in a sample aged 60+, and that occurrence of MCI varies with the applied definition. Over half of the participants with MCI reverted back to normal cognitive functioning at 6-year follow-up. Correctly assessing cognition is important and there are many methodological aspects to take into consideration, from the stage of data collection to the stage of defining cognitive impairment.

Key words Ageing, older adults, cognition, validity, test administrator effects, cohort effects, mild cognitive impairement

Classification system and/or index terms (if any)

Supplementary bibliographical information		**Language** English
ISSN and key title 1652-8220		**ISBN** 978-91-7619-798-1
Recipient's notes	**Number of pages** 93	Price
	Security classification	

Assessment of cognition in ageing. Investigating internal validity, occurrence and reversion of Mild Cognitive Impairment

Data from the general population study “Good Aging in Skåne”

Marieclaire Overton

LUND
UNIVERSITY

Coverphoto: “Brain light”, Alamy Stock Photo

Lund University, Faculty of Medicine
Department of Clinical Sciences
Doctoral Dissertation Series 2019:69

ISBN 978-91-7619-798-1
ISSN 1652-8220

Printed in Sweden by Media-Tryck, Lund University
Lund 2019

Media-Tryck is an environmentally certified and ISO 14001:2015 certified provider of printed material.
Read more about our environmental work at www.mediatryck.lu.se

MADE IN SWEDEN

We don't stop playing because we grow old, we grow old because we stop playing – George Bernard Shaw

Table of Contents

Original papers

This thesis is based on four papers referred to herein by their Roman numerals. The papers and their supplements are included at the end of this thesis. The journals publishing these papers have granted their permission for reproduction in this thesis.

I. Overton, M., Pihlsgård, M., & Elmståhl, S. (2016). Test administrator effects on cognitive performance in a longitudinal study of ageing. *Cogent Psychology*, *3*, 1260237

II. Overton, M., Pihlsgård, M., & Elmståhl, S. (2018). Up to speed: Birth cohort effects observed for speed of processing in older adults: Data from the Good Ageing in Skåne population study. *Intelligence*, *67*, 33-43

III. Overton, M., Pihlsgård, M., & Elmståhl, S. (2019). Prevalence and incidence of Mild Cognitive Impairment across subtypes, age, and sex. *Dementia and Geriatric Cognitive Disorders*

IV. Overton, M., Pihlsgård, M., & Elmståhl, S. Stability of MCI over 6- and 12-years, and factors predicting reversion to normal cognitive functioning. Data from the Good Aging in Skåne (GÅS) study. (Manuscript- planned submission)

Abbreviations

AD	Alzheimer's disease
ADL	Activity of Daily Life
APOE	Apolipoprotein E
aMCIs	Amnestic Mild Cognitive Impairment single domain
aMCIm	Amnestic Mild Cognitive Impairment multiple domain
BADL	Basic Activity of Daily Life
DSM-IV/5	Diagnostic and Statistical Manual of Mental disorders, Revised fourth edition/fifth edition
IADL	Instrumental Activity of Daily Life
GÅS	Gott Åldrande i Skåne/Good Aging in Skåne
MCI	Mild Cognitive Impairment
MMSE	Mini Mental State Examination
naMCIs	Non-amnestic Mild Cognitive Impairment single domain
naMCIm	Non-amnestic Mild Cognitive Impairment multiple domain
NCI	No Cognitive Impairment
NCD	Major/Minor Neurocognitive Disorder
TMT	Trail Making Test

Papers at a glance

Paper	I	II	III	IV
Aim	To examine the extent of potential test administrator influence on participant's test scores and other influential factors related to testing setting, such as time of day or re-test effects.	To investigate birth cohort effects on test scores measuring multiple cognitive domains in a sample of older adults as well as to examine whether level of education can in part explain the birth cohort effects.	To report prevalence and incidence of mild cognitive impairment and also investigate the how different operationalisations of cognitive impairment can influence MCI occurance estimates.	To investigate the trajectory of MCI, with specific regards to reversion rates, during a 6-year period To investigate factors associated with reversion of MCI over a 6-year period.
Population and number (N) of participants after exclusion. All are extracted from SNAC-GÅS study	Participants from 5 examination. Age *M* (*SD*) = 71.34 (10.1), 21 test administrators. N = 6686	2 samples of participants ages 60 and 81, from 3 baseline examinations. 60-year olds, N= 736 81-year olds, N = 431	Prevalence: participants from 3 baseline samples Incidence: MCI-free participants from 2 baseline samples were followed for 6-years. Prevalence, N =3752 Incidence, N = 1451	Trajectory/reversion of MCI: participants from 2 baseline samples with follow-up 6. Factors linked to reversion: baseline sample with 6-year follow-up. Participants with MCI at baseline with 6-year follow-up data, N = 331
Design	Cross-sectional and longitudinal	Cross-sectional and time-lag design	Prevalence: Cross-sectional Incidence: Longitudinal	Longitudinal cohort study
Cognitive tests	Episodic memory, speed of processing spatial ability	Speed of processing, spatial ability, memory (episodic, semantic, short-term, working), verbal fluency, attention, executive functioning, and meta cognition	Tests for cognitive impairment: Episodic memory, speed of processing, verbal fluency, spatial ability	Tests for cognitive impairment: Episodic memory, speed of processing, verbal fluency, spatial ability
Main statistical method	Linear mixed models	One-way ANOVAs (with post hoc multiple comparisons)	MCI incidence rate according to [new cases/person-years] x1000 For assessment of sex and age on incidence: Poisson regression analysis	Logistic regression model to investigate factors associated with reversion of MCI to normal cognitive functioning.
Main result	There was significant (p< 0.01) test administrator influence on participants' cognitive test scores on episodic memory, speed of processing and spatial ability.	Significant (p<0.05) birth cohort effects on participant's cognitive test scores, measuring speed of processing, episodic memory, attention, executive functioning and semantic memory (vocabulary).	Prevalence estimates ranged from 5.13-29.9% Incidence rates for overall MCI were 8.67 (95% confidence interval (CI): 7.0-10.7) and 22.6 (95% CI: 19.6-25.9) per 1000 person-years, for severe and less severe impairment, respectively.	58% of the participants reverted back to normal cognitive functioning at 6-year follow-up. Factors associated with reversion were: lower age (p<0.05), better global cognitive functioning (p<0.02), good concentration (p<0.05) and single domain subtype (p<0.001).

Introduction

The research areas ageing and cognition go hand in hand for many reasons. Among these reasons are the naturally occurring cognitive decline with advancing age (Hofer & Alwin, 2008) and that dementia, a disease primarily affecting cognition, most often onsets after the age 65 (Winblad et al., 2016). As populations around the world are growing older, there is an acute demand to fully grasp all aspects of older adult cognition.

A transition of age distribution towards older populations, referred to as *population ageing*, has long been observed. This entails that average age is higher and the proportion of older adults in population distributions is larger than it has ever been, a result of lower fertility rates and improvement of survival rates (United Nations, 2017). The World Health Organisation (2016) proposes that the proportion of people that are 60 and over will increase from 900 million (2015) to 2 billion (2050). With this ageing distribution shift, society faces a vast challenge to accommodate this growing share of seniors, as well as the increase in the occurrence of age-related diseases that affect cognition, such as dementia. Dementia is a costly disease at both the micro and macro level, bringing burden on the patient, their family and society. A recent report found a 35% increase in global dementia care cost, from US $604 billion in 2010 to $818 billion in 2015. Forecasts indicate that this cost will increase to an average global cost of $2 trillion in the year 2030 (Wimo et al., 2017). Bearing these facts in mind, given a growing population of older adults and the common occurrence of both natural and pathological cognitive decline in older populations, the world is facing a major challenge. Research focusing on ageing and cognition is therefore extremely valuable, and is the foundation of this thesis. Focus will specifically be on issues of validity when collecting and interpreting cognitive data, and on prevalence, incidence and reversion of the prodromal stage of dementia, referred to as mild cognitive impairment.

Ageing and cognition

Measuring cognitive ability

The word cognition comes from Latin *cognoscere*, meaning "to get to know" and involves mental activities that are used when obtaining and processing information (Colman, 2006). Cognition is not a unitary function responsible for all thought processes, but is further divided into various cognitive functions responsible for regulating specific behaviours. Hereinafter, these functions are referred to as cognitive domains. Notably, the taxonomy of cognition varies slightly in the literature, depending on the purpose of the research, and concepts such as cognitive domains, cognitive abilities and functions are usually used interchangeably. Domains (e.g. memory) are further divided into subdomains (e.g. short-term memory). Domains are complex, and often work simultaneously, in a synergetic manner, making it sometimes challenging to measure distinctive processes. Measurement of cognition in older adults is usually accomplished using psychometric tests or instruments specially designed to measure particular subdomains and processes, often as part of an extensive cognitive test battery. A typical test battery, measuring the cognitive status of older adults, involves tests measuring *memory, attention, language, speed of processing, executive functioning, spatial ability* and *sensorimotor skills* (Cosentino, Brickman, & Manly, 2011). The purpose of the cognitive evaluation steers who (which profession) and what type of cognitive assessment is carried out. For instance, physicians may administrate a quick global cognitive assessment such as the Mini Mental State Examination (MMSE: Folstein, Folstein, & McHugh, 1975) to gain an overall picture of cognitive level. The neuropsychologist on the other hand administrates specific neuropsychological instruments to pinpoint cognitive impairment, thus evaluating primary and secondary diagnoses, the severity and the functional boundaries of the cognitive impairment, as well as designing treatment programs (Roebuck-Spencer et al., 2017). In addition, study-trained test administrators carry out cognitive tests, often using the same standardised tests, on participants for research purposes. In order to assess level of impairment, once a test result is obtained, it is compared to a norm score, i.e. what the person is expected to score according to his/her age, education and sex (Lezak, Howieson, & Loring, 2004). In the context of ageing, evaluation of cognition can be problematic, as the examiner must determine whether the obtained test result is pathological or due to "normal" cognitive deterioration (Lezak et al., 2004).

Cognitive ageing and dementia

It is well established within ageing research that there is a natural permanent deterioration of cognition after a certain stage in maturity, referred to as cognitive ageing (Hofer & Alwin, 2008), although there is significant intra- and inter-individual variability (Wilson et al., 2002). For instance, some older adults experience more loss of cognitive functioning than others. Additionally, research has identified particular cognitive functions, e.g. vocabulary, remaining well-preserved or even improving with advancing age (Cosentino et al., 2011), whereas other functions, such as speed of processing and episodic memory, progressively decline over time (Harada, Natelson Love, & Triebel, 2013).

Besides cognitive ageing, the number one cause of cognitive decline in older adults is the syndrome dementia. Over time, dementia leads to sustained worsening of cognition, behavioural and social functioning. There are various ways of defining dementia, but generally definitions feature objective cognitive deficiency in either memory or another cognitive domain together with impairment in basic or instrumental activities of daily living. The term dementia has recently been replaced in the Diagnostic and Statistical Manual of Mental Disorders- 5 (DSM-5) with the concept *major neurocognitive disorder* (NCD) (American Psychiatric Association, 2013). This new concept embraces earlier stages of cognitive decline. Using the term NCD over dementia has its benefits, e.g. reducing stigma (Garand, Lingler, Conner, & Dew, 2009), nevertheless, dementia will be the term used in this thesis because it remains the more recognisable term. Global prevalence of dementia is 5-7% (Prince et al., 2013) and is not a single disorder but includes a range of many different, and sometimes overlapping, pathologies. The most prominent pathologies are: Alzheimer's disease, Vascular dementia, Frontotemporal dementia, Parkinson's disease dementia, Lewy body dementia, all which affect cognition in diverse ways. The umbrella image (Figure 1) provides descriptions of the primary affected cognitive domains and the prevalence estimates of the syndromes (Hogan et al., 2016; Hugo & Ganguli, 2014; Rizzi, Rosset, & Roriz-Cruz, 2014).

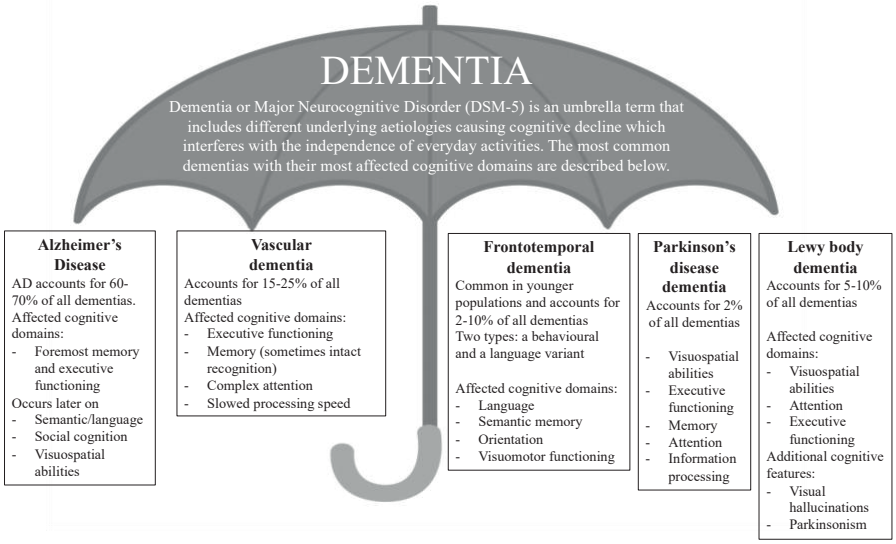

Figure 1.
Umbrella image of the most common dementia types, their prevalence and impaired cognitive domains.

Validity issues in ageing studies

Ageing research has traditionally used *longitudinal* and *cross-sectional* designs to study mechanisms of ageing. Longitudinal designs imply studying a sample of individuals over time, measuring changes in behaviour or disease outcome, whereas cross-sectional designs imply studying multiple samples of individuals simultaneously (Schaie & Willis, 2010). Within ageing research, cross-sectional studies try to answer questions regarding intraindividual age-related change using observations from interindividual differences (Schmidt & Teti, 2005), which is not optimal. To study true age-related intraindividual change, longitudinal designs are necessary. With this said, cross-sectional designs are beneficial because they can determine association between variables, and are not as costly or time consuming as longitudinal designs. There are potential validity threats in both types of study designs, as discussed in this section.

The Oxford dictionary of Psychology (2006) defines validity as:

> "The soundness or adequacy of something or the extent to which it satisfies certain standards or conditions. A research procedure or an interpretation of results obtained from a research study are considered valid if they can be justified on reasoned grounds"

Validity can further be divided into two types, *internal* and *external validity*. External validity concerns the generalisability of the conclusion of an empirical investigation. It questions whether the conclusion will remain the same when research methods or participants are changed. Internal validity concerns the accuracy of the conclusions of an empirical investigation, based on the research methods and participants that were applied. Schaie (1988) reports eight different threats to internal validity in ageing studies when using cross-sectional and longitudinal design. These are: *maturation, effects of history, practice effects, instrumentation, statistical regression, experimental morality attrition, selection bias* and finally an interaction of these. Most of these threats will be discussed in this thesis, however the next chapters will discuss instrumentation, practice effects, and effects of history more comprehensively as they are the focus of the two first papers.

Internal validity in longitudinal design

Instrumentation and practice effects, are examples of threats to internal validity when applying longitudinal designs and are discussed in the following order in the next few sections.

Instrumentation

Instrumentation refers to bias in the cognitive outcome due to changes made to the instruments measuring the intended factor. For example, participants' test results (cognitive outcome) could be biased because several versions of the same cognitive tests (instruments), intended to measure the same underlying cognitive ability, are administrated. Using multiple test administrators to collect cognitive data or slightly changing a cognitive test are example of factors that can potentially cause bias. These two types of instrumentation are fairly common in longitudinal studies as they can last for several years or even decades. Study personnel are then naturally replaced and cognitive test batteries or tests are updated (Guo & Bollen, 2013; Schaie & Willis, 2010).

Test administrator influence

As previously mentioned cognition involves a test taker (patient or a participant) and someone that administrates the test (neuropsychologist or study test administrator). When using multiple test administrators to collect cognitive data variations in testing sessions may arise due to factors attributable to the test administrators. These inconsistencies may subsequently contribute to wrongful conclusions regarding the participant's cognitive performance. This problem has been expressed in the relevant literature using the following terms: *rater agreeability, inter/intra-rater reliability* (Shrout & Fleiss, 1979), *rater effects* (Guo

& Bollen, 2013), *rater/observer bias* (Hoyt, 2000), *experimenter bias* (Rosenthal & Fode, 1963). These concepts all try to capture the extent of potential bias in cognitive outcome that arises when different observers/raters are used to evaluate the same human behaviour (Gwet, 2014). The question arises, how well do different evaluations from different rates co-inside? Notably, bias can arise when using the same observer to evaluate multiple behaviours from the same or different participants. Since testing cognition does not only consist of observational behaviour judgements, but also involves verbal interactions between the participant and administrator, the term *test administrator influence/effects* is considered more suitable and will be used in this thesis.

There are two levels of sources of variability, or bias, related to the test administrator, existing at the individual level and at the comparison level (Guo & Bollen, 2013). The first has to do with the direct impact, due to the test administrator, on cognitive performance during the testing situation. Secondly, by having several test administrators influence performance, in multiple ways, a variability in testing situations arises. By using strict and standardized procedures, both levels of bias can be substantially reduced (Rousson, Gasser, & Seifert, 2002; Sattler & Theye, 1967).

Robert Rosenthal in the 1950s was one of the first scientists to map out experimenter bias, especially pinpointing the importance of experimenter expectations on participants performance (Chapman, Benedict, & Schioth, 2018). Since then various accounts of how the administrator impacts participants' performance have been suggested. Most theoretical accounts include factors such as the characteristics of the test administrator, and the test administrator's reaction to the instrument, and/or the test taker (Chapman et al., 2018; Guo & Bollen, 2013; Sattler & Theye, 1967; Shavelson & Webb, 1991). Influence from one factor does not rule out the other, and it is most likely that test administrator influence occurs due to a combination of these factors. Instruments that are of a complex nature, and that are ambiguous in their testing techniques, instructions to participants, and scoring guidelines can cause discrepancies between test administrators (Hoyt & Kerns, 1999). Such instruments are especially vulnerable to the practice of multiple test administrators, where one test administrator adapts the correct procedure, but his/her colleague does not (Hoyt & Kerns, 1999). Ethnicity (Marx & Goff, 2005; Samuel, 1977), sex (e.g. (Chapman et al., 2018; Ortner & Vormittag, 2011; Samuel, 1977), experience of administrating (Hoyt & Kerns, 1999; Lim, 2011), and attitude (Bookout & Hosford, 1969) are examples of characteristics of the test administrator that may influence cognitive performance. An administrator's reaction to a test taker, such as expectations of performance (Rosenthal & Fode, 1963), and personal reactions to participant attributions, e.g. the Halo effect (Hoyt & Kerns, 1999; Thorndike, 1920) also assume responsibility for variations observed in test scores. A meta-analysis by Hoyt and Kerns (1999), that investigated interpretations of the rating scales and

evaluations of the target, found that up to 47% of score variance was due to the experience of the rater and difficulty level of the rating systems. The largest rater bias variance was established for inexperienced raters using inferential rating systems, i.e. they required more analytical reasoning.

Despite this evidence of test administrator influence, ageing studies that typically use multiple test administrators to collect longitudinal cognitive data fail to investigate the role of potential test administrator influence on cognitive outcome. Moreover, variation in test administration may occur more with older populations than when using younger samples, as older participants are more prone to demand extra instructions due to impairment in cognition, or/and functional issues such as impaired hearing or eyesight (Bernstein, Tucker, & Auer, 1998; Lezak et al., 2004; Sattler & Theye, 1967).

Changes to cognitive tests and batteries

As mentioned in the introduction of this chapter, cognitive tests or test batteries are occasionally replaced with updated versions. Moreover, some studies administrate a different version of the same test to the reoccurring participant at their subsequent examination in order to reduce another threat to internal validity, namely practice effects. Thus, paradoxically, by the effort to reduce one internal threat (practice effects) another one is induced (instrumentation). Regardless of the implementation reason, these versions intend to measure the same underlying cognitive mechanism. For instance, a list of words to memorise may consist of one set of words in one version and another set of words in the second version. Although these versions are constructed to be equal, there may still be inconsistencies, and one set of words may be easier to recall than the others (Katkov, Romani, & Tsodyks, 2014). For example, Laukka et al. (2013) found differences in test results between three versions measuring speed of processing tests and memory respectively. Therefore, studies applying different versions should routinely check for disparities.

Practice effects and other methodological concerns

Practice effects or *retest effects* entail an enhancement on a cognitive test performance at the successive examination due to familiarisation of that test or to the testing situation. There is evidence from longitudinal data indicating improvement in performance over examination occasions and for different types of cognitive variables (e.g. Lowe & Rabbitt, 1998; Salthouse, 2010; Wilson et al., 2002). Practice effects are especially prominent between test occasion 1 and 2 (McArdle & Woodcock, 1997) and often the effect of multiple testing decreases in magnitude across occasions (Lövdén, Ghisletta, & Lindenberger, 2004). In regards to cognitive ageing, Salthouse (2014) observed that retest effects distort the mean age trends in longitudinal comparisons when comparing three unique ageing data sets. In order to reduce practice effects in longitudinal studies, alternative forms of

the test are administrated at subsequent testing occasions (Giambra, Arenberg, Kawas, Zonderman, & Costa, 1995; Hertzog, Dixon, & Hultsch, 1992; Prince, Lewis, Bird, Blizard, & Mann, 1996). However, there is evidence that suggests that this approach does not eliminate effects of practice (Dikmen, Heaton, Grant, & Temkin, 1999; Hultsch, 1998; Watson, Pasteur, Healy, & Hughes, 1994) which signifies that other issues than learning the content of the test itself, may be contributing to improvement of test scores, e.g. *task familiarisation* or reduced anxiety (R. S. Wilson et al., 2002). Task familiarisation entails that if there are two similar tasks in one testing battery, the participant performs better on the task that is administrated secondly, due to the participant recognising how to perform this type of task (Goldberg, Harvey, Wesnes, Snyder, & Schneider, 2015). Hence, practice effects are also a concern when using cross-sectional methods. To resolve issues of task familiarity, counterbalancing task administration orders (swapping test orders) may neutralise retest effects (Ferrer & Ghisletta, 2011). Remarkably, administrating tests in different orders may cause concerns of instrumentation.

Other examples of influences on cognitive performance regarding methodological study design are the testing setting (where) and the time of day (when). That circadian rhythms affect cognitive performance is well documented, even in older adults (Blatter & Cajochen, 2007; May, Hasher, & Foong, 2005; Schmidt, Collette, Cajochen, & Peigneux, 2007). The testing setting, in which environment the person is tested, is also an influential aspect. As older adults are less mobile than younger adults, testing cognition for both research purposes, and for clinical purposes can sometimes occur outside the research centre or clinic. Previous research has found the surrounding environment to play a role on cognitive performance in older adults (Shievitz, Tudiver, Araujo, Sanghe, & Boyle, 1998; Woodford & George, 2007). Still, there is limited research conducted on influential aspects on cognitive performance when the test taker is tested in their home environment. Offering home visits to reduce internal drop-out is a vital part of keeping high participant rates, and being able to test participants at different times in the day helps keep a fast study pace. Despite the fact that these study implementations cause variations in cognitive testing, there is inadequate research conducted on the matter, especially in ageing contexts. Issues that are tied to the testing situation in our aging study Good Aging in Skåne (GÅS), such as using different test administrators or test versions, or variations in the time of day of the testing are presented in Table 1. The reason for implementing these testing specific factors and how they supposedly influence cognitive test results are also provided in the table.

Table 1.
Testing specific factors, reason for implementation and how they influence cognitive performance

Testing specific factors	Implementation reason	Latent factors
Multiple test administrators	Practical: change in personnel throughout the study	Test administrator influence, e.g. experience of test administrator
Different test versions	Reduce practice effects on subsequent testing occasions	Different difficulty levels of the test versions
Different orders of test battery administration	To reduce potential systematic fatigue effects and task familiarisation	Fatigue due to serial testing or alertness in the beginning of the testing session.
Time of day of testing	Practical: To test many participants in the same day	Fatigue due to differences in circadian rhythms.
Testing setting	To boost participant rate/minimise selection bias	Discomfort/comfort of the participant

Internal validity in cross-sectional design

Cohort effects, gains on cognitive measures

Perhaps the largest validity threat when using cross-sectional design is effects of history (Schmidt & Teti, 2005). Within ageing research, this issue involves the difficulty of separating true age-related change from the effects of history. *Period effects* and *cohort effects* are both examples of effects of history. Period effects refer to environmental circumstances that have an impact on all individuals at the same time, regardless of age or when they are born, e.g. starvation or war. Cohort effects refer to the historical effects on a group of individuals who share similar environmental circumstances at the same points in their life. A cohort could for example refer to a group of graduates or a group of individuals born during the same period, known as *birth cohorts*. These birth cohorts then share common factors such as lifestyle, cultural background, education or necessities for a successful life. These factors then contribute to the development of cognitive ability similarly, entailing that one birth cohort performs better on one cognitive test in comparison to another cohort, just by the virtue of different life experiences.

Within the field of cognition, the most prominent evidence of birth cohort effects is the substantial gain observed in intelligence scores throughout the 20th century, generally known as *Flynn effects* (Flynn, 1984; Trahan, Stuebing, Fletcher, & Hiscock, 2014). These gains are extensive and have been observed for multiple birth cohorts, age categories, as well as for multiple cognitive abilities (e.g. Hanson, Smith, & Hume, 1985; Karlsson, Thorvaldsson, Skoog, Gudmundsson, & Johansson, 2015). Gains in several tests measuring multiple cognitive functions have been reported, such as for semantic and episodic memory, verbal fluency (Rönnlund & Nilsson, 2008; Skirbekk, Stonawski, Bonsang, & Staudinger, 2013), speed of processing tasks (Karlsson et al., 2015; Salthouse, 2015; Verhaeghen, 2013), speed of processing (Dickinson & Hiscock, 2011), and spatial abilities (Karlsson et al., 2015). The magnitude of gains differs between countries and time

intervals (Pietschnig & Voracek, 2015; Williams, 2013). The general trend of gains in test scores on cognitive measures, at least in the developed countries, is currently decreasing (Flynn & Shayer, 2018; Pietschnig & Voracek, 2015). Evidence of decrease comes primarily from the Scandinavian countries, where cohort related gains are seemingly stagnating, and even reversing (Sundet, Barlaug, & Torjussen, 2004; Teasdale & Owen, 2008), but there is evidence of stagnation from other European countries (Dutton, van der Linden, & Lynn, 2016).

Causes of gains

There are many explanations of what causes the observed gains in cognition across the 20th century. For example; improved nutrition (Lynn, 2009), reduced sibling sizes (Rönnlund & Nilsson, 2008), longer and enhanced education, increased exposure to testing situations and types of tests (Brand, 1996), more experience with complex visual environment (Greenfield, 1998), and improved child rearing practices (Lynn, 2009). One explanation may not account for all gains, and certain explanations are favoured over others, which is contingent on what type of cognitive entity is proposed to be rising (Hiscock, 2007). Likewise, an interaction of different factors may explain particular gains. However, there is one explanation that is favoured over others, that is, the concurrent improvement of education (e.g. Flynn, 1984; Lynn, 2009; Tuddenham, 1948). Evidence for this account stems from the firm relationship between cognitive test performance and education (Rönnlund & Nilsson, 2008), where meta-analyses and longitudinal studies have established that education is the most persistent predictor of cognitive performance (Schaie, 2011). Furthermore, western society has observed substantial positive developments in education throughout the 20th century (Gustafsson, 2008). Evidence of education being the primary explanation for gains in cognition is persistent e.g. (Alwin & McCammon, 2001; Nyberg, Lövdén, Riklund, Lindenberger, & Bäckman, 2012; Rönnlund & Nilsson, 2008; Schaie & Willis, 2010).

Cohort effects and ageing

Several studies have reported cohort effects for older adults and a selected number of Swedish and other major ageing studies are reported in Table 2.

Table 2.
Studies reporting cohort effects using samples of older adults

Research group	Study and location	Ages	Birth years and sample size	Cognitive abilities	Outcome
Finkel, Reynolds, McArdle, and Pedersen (2007)	Swedish adoption and Twin study of ageing, Sweden	50+	Younger: 1926-1948 Older: 1900-1925 (n= 806)	Verbal and spatial ability, memory, speed of processing	Younger cohort outperformed older cohort in verbal and spatial ability and memory
Zelinski and Kennison (2007)	The long beach longitudinal study, USA	55-87	Cohort 1 1908-1940 (n=456) Cohort 2: 1893-1923 (n=482) With 16 years differences in ages between participants in the two different cohorts	Recall, reasoning, spatial ability and vocabulary	Cohort effects on all measures (recall, Recognition, reasoning, spatial ability) but not vocabulary
Rönnlund and Nilsson (2008)	The Betula study, Sweden	35-80	1909-1969 (n=2995)	Semantic and episodic memory and visuospatial ability	Advances for semantic, and episodic memory for the later born cohorts. Stagnation was noticed for individuals born 1954 and later
Skirbekk et al. (2013)	English longitudinal survey on ageing (ELSA), England	50-74	Younger: 1930-1949 Older: 1936-1955 (n=9116)	Immediate word recall, delayed word recall, verbal fluency	Younger cohorts outperformed older on all measures (immediate word recall, delayed word recall, verbal fluency)
Karlsson et al. (2015)	The gerontological and geriatric population study from Gothenburg (H70), Sweden	70, 75 and 79	Cohort 1: 1901-02 (n=460) Cohort 2: 1906-07 (n=513) Cohort 3:1930 (n=254)	Spatial ability and logical reasoning	Later born cohorts outperformed earlier born cohorts on both spatial ability and logical reasoning

Three reasons are presented below as to why cohort effects should be studied in relation to older adults. These are: 1) lack of research 2) misleading age trajectories and 3) misleading age norms.

1) Most cohort research studies use samples of younger individuals and there is currently a gap in the research regarding how cohort effects on cognitive ability are expressed in the application of older adult samples.
2) When using cross-sectional data to examine the onset of cognitive decline cohort effects may be present, causing a false observation of premature onset. Nyberg et al. (2012) demonstrate this using data from the Betula study in Figure

2a and 2b. By comparing cross-sectional data with longitudinal data, they observed a 25-year difference of when decline begins for episodic memory. However, when controlling for education levels (the primary explanation of cohort effects), the age trend for the cross-sectional data became similar to the longitudinal data, seen in Figure 2d.

3) As described in chapter 1, the determination of cognitive impairment based on test scores is accomplished through the comparisons of test scores with norm scores. These age-stratified normative scores are usually derived using cross-sectional data, and therefore norms are potentially confounded by cohort effects, leading to out-dated norms. When applying a dated norm to an individual's test score, this person could be misclassified as being cognitively intact, when in fact the person has a cognitive deficiency.

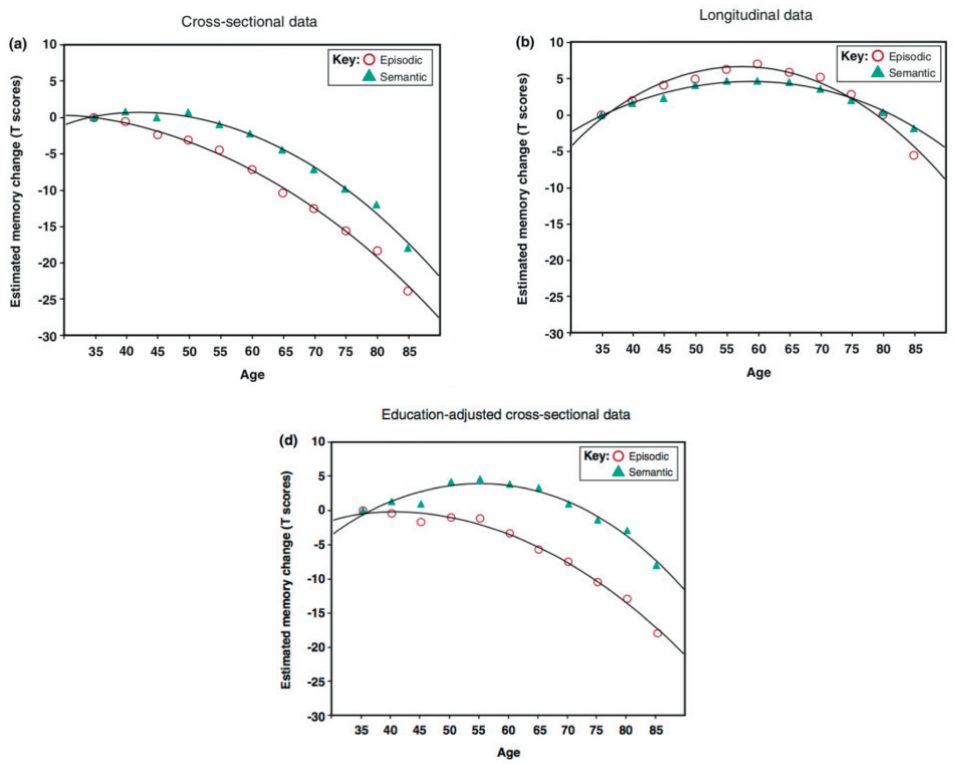

Figure 2.
Trends of age-related cognitive decline in cross-sectional (a), longitudinal (b), and education-adjusted cross-sectional data (d). Figure lent by permission (Nyberg et al., 2012).

These three reasons provide motivation to engage in research regarding birth cohort effects on cognitive performance for older adults.

In sum, it seems that by trying to reduce threats to internal validity (e.g. practice effects) in longitudinal studies by particular methodological implementations (e.g. counterbalancing test orders, or having different versions of the same tests) other validity threats are introduced. Furthermore, internal validity threats in cross-sectional studies, such as birth cohort effects, can cause inaccurate interpretation of maturation trends, as well as cause concerns when applying normative scores to determine cognitive status. Hence, ageing studies must consider the magnitude of these types of threats in their cognitive data.

Mild cognitive impairment

Dementia is one of the most detrimental age-related diseases. Thus, ageing researchers as well as clinicians are trying to understand the early stages of the disease, in order to delay its onset by introducing targeted interventions at the right time. Understanding the early stages of dementia can be done by identifying an intermediate stage between normal ageing and dementia. This transitional stage has been conceptualised in multiple ways: *Age-Associated Memory Impairment* (AAMI), *Ageing- Associated Cognitive Decline* (AACD), *Age-Related Cognitive Decline* (ARCD), *Mild Cognitive Disorder* (MCD), and *Cognitive Impairment No Dementia* (CIND)". The most recent addition to these concepts is found in DSM-5, *Mild Neurocognitive Disorder* (Mild NCD). Of these concepts, Mild Cognitive Impairment (MCI) is the definition that has received the most attention in the literature. MCI originated in the late 1980s and was further developed by Petersen and colleagues in 1999 at the Mayo Clinic in Rochester, US (Petersen et al., 1999). The MCI concept has evolved during the last few decades, steering away from its original form, which only focused on classifying those with memory impairment to ultimately capture patients with prodromal stages of AD. During a key symposium in 2003 in Stockholm (Winblad et al., 2004), the definition expanded and international criteria for MCI were formed. The definition now includes impairments in other cognitive domains than memory. Currently, the expanded Mayo Clinic characterisation of MCI includes the following features: a) self-and/or informant reported cognitive decline; b) objectively measured impairment in at least one cognitive domain; c) preserved daily functioning; and d) no dementia diagnosis (Petersen et al., 2014).

MCI is typically further separated into the four following subtypes: amnestic MCI (aMCI); single or multiple domain (aMCIs/m); non-amnestic MCI (naMCI); and single or multiple domain MCI (naMCIs/m). Amnestic MCI signifies that there is

an impaired memory component, whereas non-amnestic MCI signifies impairments in other domains. Single or multiple domain captures the number of cognitive domains affected. The purpose of the subtypes is not only to capture the pathological picture of the existing cognitive dysfunction, but to also understand its aetiology and ultimately predict dementia outcome (Petersen et al., 2014). Research has confirmed that individuals with deficiencies primarily in amnestic functions are inclined to develop AD (Belleville, Fouquet, Hudon, Zomahoun, & Croteau, 2017; Petersen et al., 2014). Furthermore, deficits in language can lead to frontotemporal dementia and visuospatial deficits may lead to dementia with Lewy bodies (Ferman et al., 2014; Nelson & O'Connor, 2008). Impairments in specific cognitive domains are measured using neuropsychological tests and the patient's/participant's test scores are therefore helpful predictors of the nature of dementia outcome. This makes the neuropsychological assessment a crucial part of the MCI definition. Markedly, additional determination of the aetiology can be made by laboratory tests or neuroimaging (Petersen et al., 2014).

Cognitive assessment of cognitive impairment

Objective cognitive impairment alone as a criterion has been found to accurately predict dementia conversion (Belleville et al., 2017; Brodaty et al., 2013; Summers & Saunders, 2012), making accurate cognitive assessment for MCI essential. Currently, there is no consensus regarding an optimal cognitive assessment battery for defining MCI and its subtypes, and test batteries described in the literature vary dramatically (Nelson & O'Connor, 2008). The existing guidelines described in Petersen et al. (2014) recommend an impairment of 1.0-1.5 standard deviations below adjusted norm scores on at least one test measuring these domains: memory, executive functioning, attention, language or visuospatial skills. This recommendation is somewhat restricted, leaving researchers and clinicians without full-scale instructions on the operationalisation of objective impairment. For instance, the number of measures for each cognitive domain, the type of psychometric measures, whether to use individual or combined cognitive measures, the quality of the comparative normative data, and the exact cut-off threshold used to indicate impairment are examples of unresolved implementation issues (Brodaty et al., 2013; Visser et al., 2009). This leads to varying operationalisations applied in the literature. Moreover, epidemiological studies and the clinical assessment in memory clinics can differ somewhat in the types of tests that are used in the cognitive assessment. For instance, most clinical neuropsychologists customize a battery of tests that suit the unique needs and abilities of the patient (Nelson & O'Connor, 2008) and also take qualitative information into consideration, whereas in epidemiological studies a standardized test battery, including multiple cognitive measures, is used for all participants (Petersen et al., 2009).

There are several other controversies and limitations in MCI research besides the implementation of the cognitive impairment criterion. For instance, source of participants (stemming from a memory clinic or population, age range), follow-up length, prospective or retrospective data collection, and classification of MCI (algorithmic or consensus approach) (Petersen et al., 2014). An algorithmic approach is usually applied in the context of population-based studies, where criteria for MCI is listed and diagnosis is assumed only if each of the criteria is fulfilled. Mostly, focus lies on whether the participant is considered to have an impaired neuropsychological test score or not. A consensus approach on the other hand is when an expert panel is provided with relevant clinical information, such as life span experience or concerns from relatives to make a justified decision of MCI diagnosis. These controversies lead to methodological differences in MCI research, which contribute to heterogeneous research findings on MCI regarding prevalence, incidence, progression to dementia, risk markers, sex and age differences, and stability of the diagnosis.

Although beyond the scope of this thesis, it is notable that the diagnosis of MCI extends further than cognitive examination. Medical history, functional status, inquiry by proxy and collection of biomarkers are vital factors to make an accurate clinical judgement of preclinical dementia. Together with the abovementioned controversies, major differences in the inclusion and the weight of these factors exist, further complicating comparisons of research results.

Prevalence estimates of MCI

An average MCI prevalence of 18.9% has been put forward for individuals over 65 (Petersen et al., 2014). A recent meta-analysis by Petersen et al. (2018) and colleagues reported prevalence for population-based MCI between 3.2% - 42.0% (for both overall MCI and amnestic MCI), which is comparable to the prevalence reported in another meta-analysis (3-42%) conducted by Ward, Arrighi, Michels, and Cedarbaum (2012). Similarity between meta-analyses is explained, in part, by overlapping studies.

MCI generally increases with age (Petersen et al., 2018), a notion that has been supported in multiple research studies (Manly et al., 2005; Ravaglia et al., 2008; Roberts et al., 2014; Solfrizzi et al., 2004; Trittschuh et al., 2011). Petersen et al. (2018) reported the following prevalence estimates for age groups 60-64: 6.7%, 65-69: 8.4%, 70-74: 10.1%, 75-79: 14.8%, 80-84; 25.2%.

Regarding sex-differences in MCI estimates, the literature provides an ambiguous picture, with some studies finding higher overall MCI prevalence for women (Artero et al., 2008; Lee et al., 2009) and some for men (Annweiler et al., 2012; Guaita et al., 2015; Petersen et al., 2010). A meta-analysis conducted by Au, Dale-

McGrath, and Tierney (2017) that included 56 studies, reported no sex-differences for overall MCI prevalence, but observed a significantly higher prevalence among women for non-amnestic MCI.

Incidence estimates of MCI

The average incidence rate of MCI (using the expanded Mayo Clinic criteria) for individuals over 65 was reported in Ward et al. (2012) to be 47.9 per 1000-person years (range: 21.5-71.3), and rates for aMCI ranged from 8.5 to 25.9 per 1000-person years. Higher incidence rates have however been reported by Luck, Luppa, Briel, and Riedel-Heller (2010): 76.5 (CI: 64.7-90.4) and Brodaty et al. (2013): 104.6 per 1000-person years (CI:81.6-127.7).

Incidence rates are reported to increase with age in some studies (Luck et al., 2010; Roberts et al., 2014; Solfrizzi et al., 2004) and others report no age effects on incidence (Brodaty et al., 2013; Busse, Hensel, Gühne, Angermeyer, & Riedel-Heller, 2006; Larrieu et al., 2002; Ravaglia et al., 2008). A very recent meta-analysis reported incidence per 1000 person-years (95% CI) for the following age-categories: 22.5 (5.1– 51.4) for 75–79 years, 40.9 (7.7–97.5) for 80–84 years, and 60.1 (6.7–159.0) for 85 + years (Gillis, Mirzaei, Potashman, Ikram, & Maserejian, 2019). Regarding sex-differences in MCI incidence, some studies have found higher incidence for men (Brodaty et al., 2013; Ravaglia et al., 2008; Roberts et al., 2014) and some for women (Luck et al., 2010; Tervo et al., 2004). The meta-analysis by Au et al. (2017) detected no sex-differences for amnestic, non-amnestic or overall MCI. However, the authors did observe higher incidence rate ratios for men than women in all the examined studies.

Variation in estimates of MCI occurrence due to the applied operationalisation

Depending on the operationalisation applied to define MCI prevalence, estimates have been found to vary between 11% and 92% (Trittschuh et al., 2011) and incidence rates range between 23.5 and 104.6 per 1000 person-years (Brodaty et al., 2013). One operationalisation principle involves deciding how many impaired test scores are required for the classification of objective cognitive impairment. MCI studies differ in this respect, leading to disparities in prevalence and incidence estimates reported in the literature. For instance, the Mayo Clinic definition only requires one impaired test score, whereas Jak/Bondi require two (Bondi et al., 2014). A prevalence of 24% was reported using the Jak/Bondi criteria whereas when using the Mayo Clinic criteria prevalence was 34% (Jak et al., 2016). A substantial difference in incidence rates for one vs. two impaired test scores was shown in Brodaty et al. (2013), 1 test: 104.6 and 2 tests: 23.5 per 1000-person years. Hence,

more precise guidelines are warranted for the application of cognitive tests in MCI, especially in relation to the number of tests to determine impairment.

MCI trajectory and reversion to normal cognition

MCI can follow different evolutionary trajectories (Diaz-Mardomingo, Garcia-Herranz, Rodriguez-Fernandez, Venero, & Peraita, 2017).

a) *Progressive* - over time the MCI diagnosis progresses to dementia

b) *Stable MCI* - over time the MCI diagnosis remains the same, with impairment in the same cognitive domain or changing domains

c) *Unstable* - over time the MCI diagnosis fluctuates, going back and forth between normal cognitive functioning and cognitive impairment

d) *Reversion* - recovering from MCI and being considered cognitively intact

See Figure 3 for a visual description of MCI trajectories.

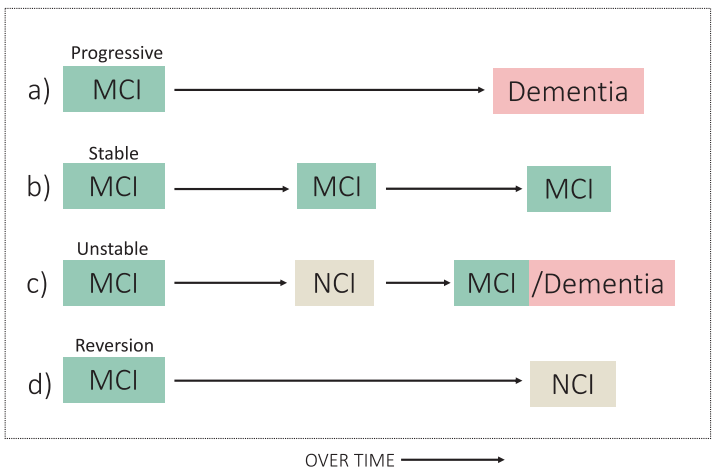

Figure 3.
This is an overview of four possible trajectories of MCI. Abbreviations: MCI = Mild Cognitive Impairment, NCI = No Cognitive Impairment

The first alternative (a) represents the primary purpose of the MCI concept and annual rates of progression from MCI to dementia (all types) is reported to be 6-15%. The magnitude of the progression rate is partially dependent on whether the

sample is from a population or a memory clinic (Petersen, 2011). For example, a meta-analysis examining 41 studies found an annual progression rate of 10% in clinical samples and 5% in community samples (Mitchell & Shiri-Feshki, 2009). In addition, Petersen et al. (2018), in a meta-analysis of nine high quality population-based articles, described an increased risk of conversion to dementia for MCI participants in comparison to MCI-free and age-matched participants. In the same article, cumulative progression rates from MCI to dementia over 2 years for individuals 65+ was 14.9% (CI:11.6%-19.1%). The size of the annual progression rate suggests that the MCI trajectory is no straight path to dementia, with the majority of individuals either remaining stable, unstable, or reverting back to normal cognitive status.

There is evidence suggesting that the smallest group progresses to dementia and the largest remains stable MCI (b in Figure) (Pandya, Clem, Silva, & Woon, 2016). Stability estimates using the expanded Mayo Clinic MCI definition from both European, Asian, Australian and American population-based studies are reported to be between 37%-67% (Pandya et al., 2016). Interestingly, of those who remain stable MCI over time, some of these individuals switch between MCI subtypes (e.g. from amnestic MCI to non-amnestic MCI) (Aerts et al., 2017; Han et al., 2012). The two remaining MCI trajectory groups are the unstable group and the reversion group. The unstable group (c in Figure) is a remarkable group, as they revert back to normal cognition and then retransition to MCI or progress to dementia. Few studies have described this unstable trajectory. Lopez et al. (2012) found that 10.5% of their sample from the Cardiovascular Health Study-Cognition Study had an unstable MCI trajectory. Similarly, Roberts et al. (2014) reported that 65% of those that reverted back to normal cognitive functioning re-transitioned to MCI at follow-up. This indicates that it is fairly common to fluctuate in MCI diagnosis. Moreover, this unstable group has a higher risk of progressing to dementia in comparison to those who have never had an MCI diagnosis (Lopez et al., 2012; Roberts et al., 2014).

The last alternative (d in Figure) concerns reversion from MCI to normal cognitive status at follow-up. Reversion rates of 29% (Canevelli et al., 2016) and 26.4% (Petersen et al., 2018) for overall MCI in population-based samples have been documented, ranging from 4-58%. For amnestic MCI Malek-Ahmadi (2016) discovered that 31% reverted back when examining 17 population-based studies.

Despite the reversion group being the second largest group among alternative trajectories for MCI, few studies have had the primary focus on reversion and factors associated with reversion (Canevelli et al., 2016; Malek-Ahmadi, 2016; Pandya et al., 2016). Moreover, clinicians need a diagnostic tool that can accurately identify individuals at a prodromal stage of dementia but can also identify individuals with

a favourable prognosis. The coming sections will discuss factors associated with reversion.

Factors tied to reversion

In comparison to predictors (risk factors) of dementia progression, very few studies have focused on predictors of MCI reversion. The first study that actually focused on factors associated with reversion was conducted by Koepsell and Monsell (2012) using a sample from the National Alzheimer's Coordinating Center, aged 65 and over, with a follow-up of one year. Results showed that 20% progressed to dementia, 64% had stable MCI and 16% reverted. They found that factors associated with reversion were: younger age, absence of self-and/or informant reported cognitive decline, higher global functioning, lower scores on The Clinical Dementia Rating Scale Sum of Boxes and lower scores on The Functional Activities Questionnaire, non-amnestic single domain subtype, and no Apolipoprotein ε4 allele (APOE ε4 allele). The abovementioned reversion factors have further been confirmed, and new predictors have been exposed when reverters are compared to non-reverters or dementia progressors.

Person based factors, such as physical health status or demographics have been found to predict reversion. For example, younger age, the male sex, higher and lower education and having life partner as well as being single are examples of person-based factors tied to reversion (Aerts et al., 2017; Han et al., 2012; Pandya, Lacritz, Weiner, Deschner, & Woon, 2017; Roberts et al., 2014; Sachdev et al., 2013; Tokuchi et al., 2014). In addition, good global cognitive functioning (e.g. higher scores on MMSE), better performance on neuropsychological test, improvement of depression, openness to experience as a personality trait, and better smelling/vision ability have also been seen to be associated with reversion (Han et al., 2012; Koepsell & Monsell, 2012; Pandya et al., 2017; Park et al., 2015; Roberts et al., 2014; Sachdev et al., 2013; Sugarman, Alosco, Tripodis, Steinberg, & Stern, 2018). Biological factors associated with reversion are larger drop in diastolic blood pressure (Sachdev et al., 2013), greater hippocampal volume (Park et al., 2015; Sachdev et al., 2013) or lack of the APOE ε4 allele (Pandya et al., 2017; Park et al., 2015; Roberts et al., 2014).

Other predictive factors that are more related to the diagnostic criteria applied have also been suggested. For instance, absence of self or informant complaint and using liberal MCI criteria to classify MCI (Sachdev et al., 2013), single domain (Han et al., 2012; Roberts et al., 2014; Sachdev et al., 2013), and non-amnestic MCI (Roberts et al., 2014). In addition, the misdiagnosis of MCI caused by the usage of inappropriate normative scoring, e.g. *bracket creep*: moving into a higher age group with lower norms, has been proposed (Aerts et al., 2017), but not yet studied.

Aims

The overall aims of the studies were

Paper I

To inspect to what extent variations in testing setting factors, such as using multiple test administrators to collect data, could influence participants' cognitive test performance.

Paper II

To explore birth cohort effects on test performance that measure multiple cognitive domains and examine whether level of education could explain the observed effects.

Paper III

To report prevalence and incidence of mild cognitive impairment across age, sex and subtypes and investigate how different operationalisations of cognitive impairment could influence MCI estimates.

Paper IV

To investigate the trajectory and reversion of mild cognitive impairment using a 6- and 12-year period and assess factors for reversion from MCI to normal cognitive functioning.

Methods

The Good Aging in Skåne study

Participants for all four papers were sampled through the Good Aging in Skåne study (GÅS: Gott Åldrande i Skåne), which is an ongoing longitudinal study in the southern part of Sweden, covering five municipalities including both rural and urban areas (Ekström & Elmståhl, 2006). GÅS is one of four participating sites in the Swedish National Study on Aging and Care (SNAC) (Lagergren et al., 2004). In 2001 (Wave 1), men and women aged 60, 66, 72, 78, 81, 84, 87, 90, 93, 96, 99 were invited by letter and at random from the general population using the National Population Registry. Non-respondents were contacted again by letter or telephone, and those unable to come to the test centre were primarily offered home visits and secondarily telephone interviews. Participants younger than 78 years of age were invited back for re-examination every six years and for participants older than 78 every three years. In addition, new participants aged 60 and 81 were recruited in 2007 (Wave 2), 2012 (Wave 3) and in 2018 (Wave 4) and the last wave is ongoing. Each examination took 2-3 years to complete, and so far, examination Wave 1 has five completed follow-up examinations, with the sixth ongoing. Wave 2 has four follow-up examinations, with 12-year examination data currently being collected and Wave 3 has two follow-up examinations with 6-year examination data being collected. See Figure 4 for study flow.

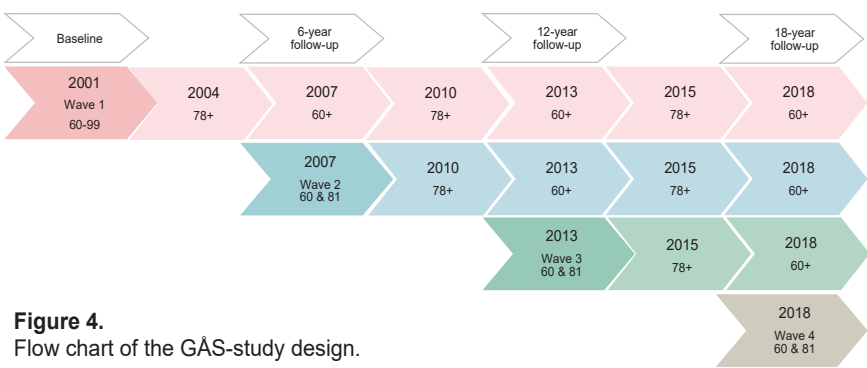

Figure 4.
Flow chart of the GÅS-study design.

This entails that the GÅS study (in year 2019) has three completed examination waves with baseline data. The participant rate for Wave 1 was 60%, and for both Wave 2 and 3 the participant rate was 70%. During a typical full-day of examination at one of the five test centres, a participant met four professionals: a doctor/physician, a registered nurse, a study-trained psychological test administrator and a medical secretary. The physician carried out an extensive health examination including medical history and Electrocardiography and the nurse performed functional tests, spirometry, laboratory tests, and measured anthropometrics. Participants completed a comprehensive psychological examination during the test administrator examination, including a cognitive test battery and interviews. The medical secretary administrated questionnaires regarding current and former life style, social circumstances, day-to-day functioning, diseases, and medications. A full day of examination took about seven hours to complete for each participant.

Study samples

All samples for papers I, II, III and IV are taken from the GÅS study and flow charts describing study selection are presented below. The joint exclusion criteria for all papers was to exclude participants:

- with dementia
- who did not complete a cognitive testing
- with hearing and vision issues that influenced how well they performed on the cognitive tests
- who did not have Swedish as their native tongue and who were therefore at risk of performing worse on Swedish-language tests than native Swedish speakers

Paper I: Test administrator effects

Participants included in paper I came from five examination times, including both participants with baseline data (stemming from Waves 1, 2 and 3) and follow-up data (re-examined participants from Waves 1 and 2). Because one participant can exist multiple times in this sample, each data point is referred to as participant observation rather than individual participants. See Figure 5.

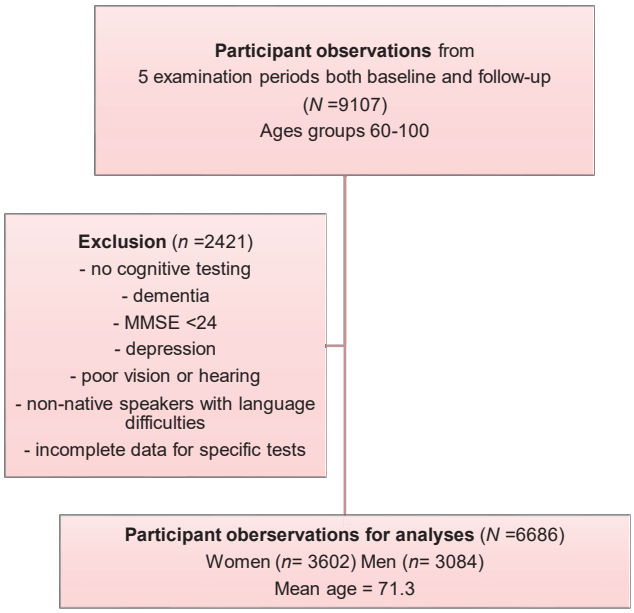

Figure 5.
Flow chart for participant selection in paper I

Paper II: Birth cohort effects

Inclusion for paper II was participants the same ages, when examined, but born 5-7 years apart. Every sixth year, new participants, aged 60 and 81 are invited to partake in the GÅS-study, leading to three baseline data waves so far. Participants aged 60 were born: 1942-43, 1948-49 and 1954-55. Participants aged 81 were born: 1920-21, 1926-27 and 1932-33. See Figure 6.

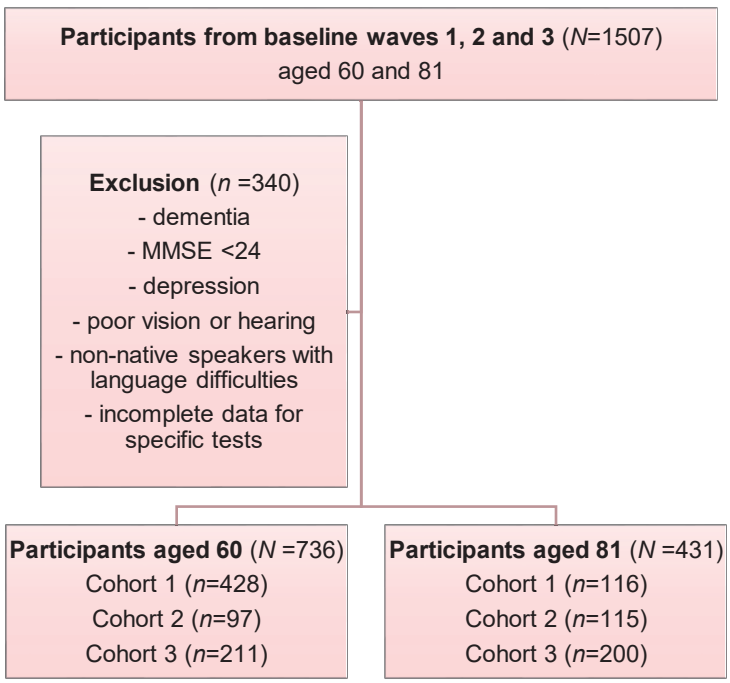

Figure 6.
Flow chart for participant selection in paper II

Paper III: Prevalence and incidence of MCI

Participants for MCI prevalence calculations in paper III were taken from baseline Waves 1, 2 and 3. Wave 2 and 3 had 6-year follow-up data and therefore data for incidence calculations were taken from these two waves. No differences in MCI prevalence estimates were identified between the examination waves, endorsing the decision to merge datasets. See Figure 7.

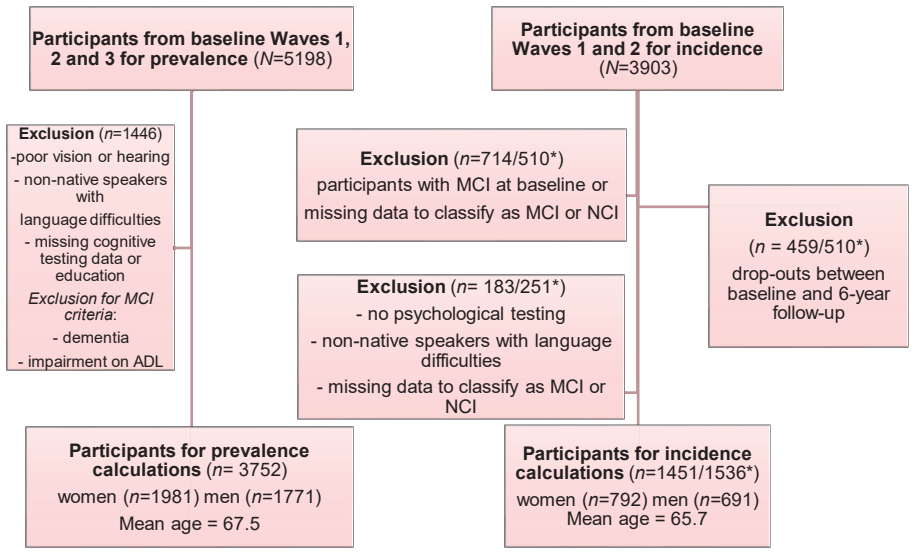

Figure 7.
Flow chart for participant selection in paper III. *lenient/strict criteria for MCI. The numbers of participants vary with the criterion used to define MCI. Lenient criterion = at least one impaired test score in a cognitive domain. Strict criterion = at least two impaired test scores in each cognitive domain. See section "defining MCI" for further details.

Paper IV Trajectory of MCI

Participants for paper IV were selected from Waves 1 and 2 with 6-year follow-up data to examine the MCI trajectory and factors associated with reversion. Inclusion criterion was an existing MCI diagnosis at baseline. See Figure 8. For 12-year trajectory, participants from examination Wave 1 (n=212) were selected with a 6-year follow-up examination and a 12-year follow-up examination (no flow chart).

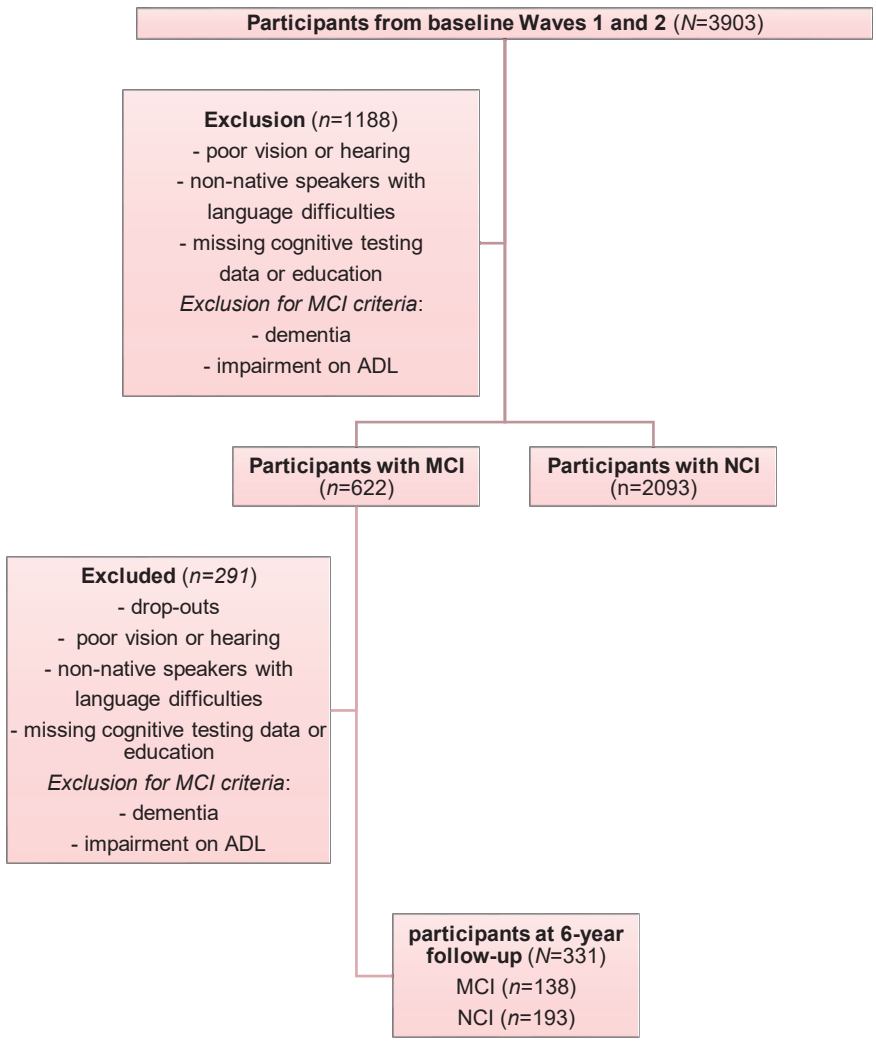

Figure 8.
Flow chart for participant selection in paper IV

Cognitive testing procedure

The cognitive testing session was carried out by a test administrator and comprised of 12 cognitive tests, tailored for an older population. After the tests were administrated, participants were interviewed regarding major life events, depression, and coping strategies. The examination took roughly 1.5 hours to complete. A structural equation model conducted by GÅS's sister study SNAC-Kungsholmen identified the following latent cognitive domains (Laukka et al., 2013): episodic memory, speed of processing, semantic memory, language (word fluency letters and categories). The remaining tests were assigned to the following cognitive domains: visuospatial ability, attention, executive functioning (Pantzar et al., 2014), short-term memory, working memory (Gignac, 2015) and metacognition (Dahl, Allwood, Rennemark, & Hagberg, 2010).

Test administrators

At current there have been 25 test administrators involved in the GÅS-study. However, at the point of when paper I was conducted, there had been 21 test administrators working with data collection, 2 men, 19 women, with a mean age of 32.8 (range: 25-50) and a median age of 30. The average duration of employment was 2.19 years, ranging from a few months to 10 years. The minimum requirement for employment was a bachelor degree in behavioural sciences or the equivalent. The test administrators were trained by an experienced test administrator for at least two weeks, and were provided with a detailed instruction handbook. Prior to testing his/her first participant, each test administrator directed a testing session with a mock participant. Routine check-ups of the test administrators' execution of cognitive test battery, together with frequent meetings with a neuropsychologist and other test administrator have been implemented. This was to ensure standardised administration and consistency in test scoring. To minimise expectancy effects, the test administrators had no knowledge on how well the participant performed on previous cognitive tests.

Time of day, test order and test version

One of three examination slots during the day was unsystematically allocated to the participant, between 8 and 10 am, or prior to lunch, between 10 am and noon or in the afternoon, between 1 and 3 pm. Systematic effects of fatigue due to serial testing were minimized through the practice of administrating the cognitive testing battery in two different orders. Also, to minimize practice effects, three different versions for each test were provided when possible, with the participant receiving a different version at their follow-up visit. The three versions were constructed to measure the same cognitive ability and to be equally difficult, however slight differences

between the tests have been identified (Laukka et al., 2013). Both test orders and test versions were divided equally between all participants.

Cognitive tests

Episodic memory

Episodic memory was measured with a free recall and recognition test in all four papers. The test consisted of learning a word list composed of 16 unrelated concrete nouns (target words), with an immediate free recall trial and a forced yes/no recognition trial (Gardiner & Java, 1993). Words were presented one by one, both orally and visually, in a booklet for five seconds. Participants were requested immediately after the presentation to remember as many words as possible. For the self-paced recognition task, the 16 target words intermixed with 16 interference words and were presented in a booklet. Test assessment for free recall was the total number of correctly recalled words, with a maximum score of 16. For paper I, *d*-prime was used for test score assessment. For papers II, III and IV recognised words (hits) minus false hits were used (Laukka et al., 2013). A five-item recall task (Hagberg, Bauer Alfredson, Poon, & Homma, 2001) was also used to assess episodic memory in papers III and IV. During the medical examination five common items (e.g. key or paper clip) were placed in front of the participant, and he/she was requested to remember the items after 10-15 minutes. Three or less recalled items was considered a pathological score (Hagberg et al., 2001).

Speed of processing

To measure speed of processing in all four papers, the pattern comparison task (Salthouse & Babcock, 1991) and the digit cancellation task (Zazzo, 1974) were used. The pattern comparison task consisted of two separately administered pages, with 30 pairs of line-segment patterns, half of which were identical and half different, divided equally into two columns on each page. The participant was asked to identify as many identical line-segments as possible during 30 seconds for each page respectively. The digit cancellation task consisted of one page with 11 rows of digits, ranging from 1 to 9, with one isolated practice row at the top of the page. There were 43 ‘4’s and the participant had 30 seconds to cross over as many of these as possible. Test score assessment for pattern comparison was the total number of correctly classified patterns from two pages divided by two, and for digit cancellation was the total number of correctly crossed over digit ‘4s.

Semantic memory

In paper II, a vocabulary test (Dureman, 1960) and a general knowledge test (Dahl et al., 2010) were used to measure semantic memory. For the vocabulary test,

participants were requested to select synonyms for 30 words from five alternatives for each word. Test score assessment was the total number of correct synonyms during seven minutes. The general knowledge test was comprised of ten general knowledge questions (e.g. What is the capital of Uruguay?), with a multiple choice of two answers. Test score assessment was the total number of correct answers.

Language

Language in papers II, III and IV was measured with two verbal fluency tests (described in Strauss, Sherman, & Spreen, 2006; Tombaugh, Kozak, & Rees, 1999): letters (phonemic fluency) and categories (semantic fluency). For letter fluency, participants were requested to generate as many words as possible beginning with the letter A and then F, during 60 seconds respectively. For the categorical word fluency, participants were required to produce as many animals/work professions as possible during 60 seconds. Test assessment was the total number of words from the two tasks (A&F) or (animals and professions) divided by two.

Attention and executive functioning

A shortened version of the trail making A (TMT-A) created for SNAC (Pantzar et al., 2014; Reitan & Wolfson, 1985) was used to measure attention in paper II. TMT-A consisted of a piece of paper with 13 circles, where each circle had a number in it. The participant's task was to draw a line between the circles in ascending numerical order (e.g. 1-2-3) as quickly as possible. A shortened version of the trail making test B (TMT-B) created for SNAC was used to measure executive functioning in paper II. Here again, 13 circles were presented on a piece of paper, this time with numbers and letters in the circles. The participant's task was to connect the circles in alternating numerical and alphabetical order (e.g. 1-A-2-B) as quickly as possible. Test score assessment was the completion time measured in seconds for the 12 correct connections.

Short-term memory and working memory

Short-term memory and working memory in paper II were measured with digit span forwards and digit span backwards, respectively. The task digit span used in GÅS was based on the digit span tests described in Manual for the Wechsler Adult Intelligence Scale—Revised (Wechsler, 1981). For digit span forward, participants were requested to repeat a series of single digits in the correct order after the test administrator presented them orally to the participant (e.g. 1-2-3). Digit span backwards was the same procedure, but the participant was to recall the digits in the reverse order (e.g. 1-2-3 and response 3-2-1). Test score assessment for digit span forwards and backwards was the total number of recalled digits (minimum 2 and maximum of 9 digits) in the last correctly recalled series (minimum 2 and maximum 7 series).

Metacognition

An accompanying task to the general knowledge test mentioned previously was used to measure metacognition in paper II. This task requested the participants to make confidence judgements for their answers on the ten questions in the general knowledge test. For each question, the participants specified how confident they were that their answer was correct by selecting a percentage between 50-100. The selection of 50% meant that they had completely guessed and 100% meant that they were 100 percent certain that their answer was correct. Test score assessment was a calibration formula by Dahl et al. (2010), which takes both the total number of correct answers and the participant's confidence judgements into consideration. Dahl et al. (2010) reported a high internal consistency for this task (Cronbach's alpha of 0.78).

Visuospatial ability

Visuospatial ability was measured using a mental rotation test (Rehnman & Herlitz, 2006) in papers I, II, III and IV. Figures of the Shepard and Metzler type (Shepard & Metzler, 1971) were presented in a booklet. There were ten series of four two-dimensional figures. Each series comprised one original figure positioned to the left of three figures and one of the figures to the right was the same as the original figure but rotated. The participant's task was to correctly identify the rotated figure to the right as quickly as possible. Two aspects of this task were used for test score assessment: speed and accuracy. Speed was measured by the total number of seconds it took for the participant to point at the correct figure when controlling for the difficulty level of that specific rotation series. Accuracy was measured by the total number of correct answers (maximum 10), divided by the number of series the participant completed (maximum 10). Most participants completed all 10 tasks.

Depression

At the end of the testing session, the Comprehensive Psychiatric Rating Scale (CPRS) was used to assess the participant's levels of depression (Montgomery & Asberg, 1979). Ten questions, with a scale from 0-6, were used to assess aspects of depression such as: sadness, anxiety, concentration and sleep deprivation. Scores between; 35-60 were considered severe depression, 20-34 moderate depression, 7-19 mild depression and under 7 no depression (Snaith, Harrop, Newby, & Teale, 1986). The CPRS has repeatedly been used to assess depression in elderly populations (Bäckman, Hill, & Forsell, 1996; Pantzar et al., 2014). In papers I and II moderate and severely depressed participants were excluded. In papers III depression was used as a descriptive factor. In paper IV, depression was used as a predictive factor for MCI reversion.

MMSE

The mini-mental state examination (MMSE) (Folstein et al., 1975) is one of the most popular cognitive screening instruments used and is commonly used as a measure of global functioning. Participants are requested to perform tasks measuring orientation, attention, recall, language and spatial ability, with a maximum possible score of 30. Participants with MMSE scores of <24 (Tombaugh & McIntyre, 1992) were excluded in papers I and II. In paper IV MMSE-scores were used as a predictive factor for reversion.

Concentration and motivation

The level of concentration and motivation of the participant when performing the tests was assessed qualitatively by the test administrator. The participant was assessed to either have good, moderate or bad concentration/motivation. In Paper IV moderate and bad concentration were merged into one factor characterising bad concentration.

Other measurable concepts

Education

The participant's level of education was determined using two methods. In papers I and II, education was defined by the total number of years the participant had attended school, school included all formal education. This information was obtained during the test administrator interview. In papers II, III and IV education were defined categorically, with three levels of education: primary school, secondary school and university degree.

Dementia

Dementia was diagnosed by the examining physician according to the Diagnostic and Statistical Manual of Mental Disorders-IV criteria. This definition includes the development of several cognitive deficiencies manifest by: (1) impairment in memory and (2) at least one of the following cognitive disturbances: (a) aphasia (language disturbance) (b) apraxia (impaired capability to carry out motor activities despite intact motor function) (c) agnosia (inability to identify objects despite undamaged sensory function) (d) disturbance in executive functioning (i.e., planning, organising or the ability to use abstract reasoning) and (3) that the cognitive deficiencies significantly impair the persons social or occupational functioning and signify a substantial deterioration from the previous level of functioning. Evaluation of deficits in memory or other cognitive areas are based on: results on MMSE, the five-item test, Crooks screening scale (Crook et al., 1986)

and the Clinical dementia Rating (CDR) (Hughes, Berg, Danziger, Coben, & Martin, 1982). Dementia diagnosis is also based on the patient's medical records.

ADL

A revised version of Katz' Activities of daily living (ADL)-index (Katz & Akpom, 1976) was used to assess participants' functional abilities in papers III and IV. This self-rated instrument was presented in a questionnaire and comprised 10 questions measuring two functional domains, Basic Activities of Daily Living (BADL) and Instrumental Activities of Daily Living (IADL). BADL involves questions regarding self-maintenance functions such moving to and from the bed, eating self-reliantly, getting dressed and undressed, using the bathroom/toilet and having issues with incontinence. Whereas, IADL involves questions concerning grocery shopping, cooking food, heavier household chores and transportation. If the participant requires assistance when performing one of these daily functions then the participant is considered dependent in their daily activities. In papers III and IV, participant's with dependency in BADL and IADL were excluded from the MCI groups. A minimal dependency in IADL (a score of 0-1 out of 4 questions on IADL) was considered acceptable in the MCI group as it is common for persons with mild MCI to present problems with IADL (Petersen et al., 2014).

Defining MCI

Mild cognitive impairment in papers III and IV was defined based on the expanded original Mayo Clinic criteria (Petersen et al., 2014): 1) *subjective and/or informant cognitive* complaint was defined by the participants themselves or by the examining physician; 2) *normal functional ability* was defined by the Katz's ADL scale mentioned above; 3) *absence of dementia* defined by the examining physician using DSM-IV and medical records; and 4) *objective cognitive impairment* was defined was impairment in one or more cognitive domains relative to normative data. The number of cognitive tests in which cognitive domain that were used to define cognitive impairment are found in Table 3. Normative scores for the cognitive tests were generated using a sample from the GÅS sister study in Blekinge (SNAC-Blekinge) and 20% of randomly selected participants from the GÅS baseline sample (Wave 1). The two samples were comparable concerning test scores in the particular cognitive domains, which provided reassurance regarding the validity of the decision to merge of the two samples. A test score below the 7th percentile, corrected for age, sex and education, was considered an impaired test score. When assuming distribution normality, the 7th percentile equates to 1.5 standard deviations below the mean.

Table 3.
Number (n) of tests used to measure each cognitive domain and the number of tests required for the lenient or strict criterion

Domain (number of tests)	Lenient	Strict
Memory (3)	1	2
Speed of processing (2)	1	2
Verbal (2)	1	2
Spatial (1)	1	2

Note. If 2 or more test scores in each relevant domain are impaired then the case is considered strict MCI. If at least 1 test score in each relevant domain is impaired, then the case is considered lenient

Further, four subtypes depending on number and type of cognitive impairment were applied: Amnestic MCI single domain (aMCIs), Amnestic MCI multiple domain (aMCIm), non-amnestic MCI single domain (naMCIs) and non-amnestic MCI multiple domain (naMCIm). See Figure 9 for implementation of MCI in papers III and IV. In paper III, two criteria regarding severity of impairment were applied, a lenient and a strict criterion. For the lenient criterion at least one impaired test score in a cognitive domain was required and for the strict criterion the requirement was at least two impaired test scores. In paper III and IV, the lenient/strict criteria were used to quantify less severe/severe MCI.

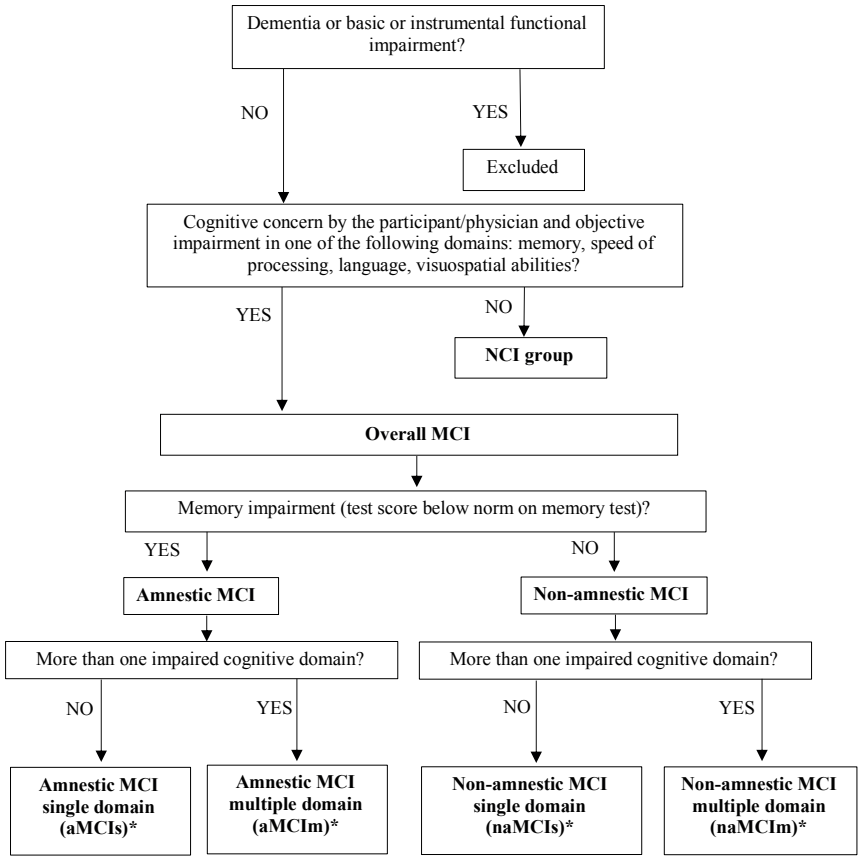

Figure 9.
Note. Abbreviations: NCI: No cognitive impairment. NCI: all scores are over normal cut-off, for participants with missing scores on tests they have to score < 27 on the mini-mental state examination to classify as healthy (Tombaugh & McIntyre, 1992). *The participant is additionally divided into lenient or strict criteria of the MCI subtype. Lenient criteria: at least one impaired test score in each relevant domain. Strict: at least two impaired test score in each relevant domain.

Ethical considerations

The studies included in this thesis are all part of the GÅS-project which has been approved by the regional Ethics committee of Lund University, Sweden. The registration number is LU 744-00 and the approval was conducted following the guidelines of Declaration of Helsinki developed by the World Medical Association.

All participants provided written consent and were informed that they had the right to withdraw at any time. The anonymity was ensured through the application of

individual serial numbers for each participant. This serial number could not be traced back to the person by the researcher.

GÅS-participants are exposed to a small amount of risk of psychological strains. The participant is namely asked personal questions about depressive symptoms and negative major life events which may provoke sadness. Moreover, participants may react on the personal inquisition, and not want to discuss personal experiences with someone they do not know. The research personnel are trained in dealing with these delicate situations, where they assure the participant full anonymity and that their answers are not traceable back to their names, that they can terminate the examination at any time and that they do not have to share anything they do not want to. Often, this is enough to sooth the situation and make the participant feel comfortable. What is more, to perform cognitive tests could be a hurtful psychological experience, where the participant experiences that they did not perform well on the tests and realise that their mental abilities are not what they used to be. The benefit of participating is then that if there is a mental deterioration then the participant is remitted to the memory clinic or their general practitioner for further examination.

Statistical methods

Table 4 demonstrates the statistical procedures applied in the four papers together with outcome variables and predictor variables.

In paper I, a series of mixed linear models' analyses were performed, as we judged that this form of analysis was necessary given the structure of the data and that our primary purpose was to investigate whether there was variance in participants' test scores attributable to the test administrator. Mixed models should be used when dependency in the data is present. Participants' test scores may not be independent when guided by the same test administrator. Similarly, the scores coming from the same participant at different occasions may be dependent. Therefore, 'test administrator' and 'participant' were included as random effects. We used restricted maximum likelihood estimation to derive parameters as our main focus was on producing accurate estimates of random variances rather than using the maximum likelihood that produces more accurate estimations for fixed effects (Twisk, 2006).

In paper II, differences in test results across birth cohort were investigated using one-way ANOVAs. Pairwise differences between birth cohorts were analysed using post-hoc multiple comparisons and linearity of cohort trends was investigated using suitable contrast tests. To investigate the role and impact of education on cohort effects, two (cohort 1 vs. cohort 2 and cohort 1 vs. cohort 3, respectively) one-way ANOVAs without education were run, followed by two 2-way ANOVAs with

adjustment for education categorically and lastly two 2-way ANCOVAs were run adjusting for education continuously.

In paper III, crude incidence was calculated using the following formula 1000[cases/person-years]. To investigate effects of age and sex on incidence rates, a Poisson regression analysis was chosen.

In paper IV a series of logistic regressions were run to examine factors related to MCI reversion.

Normative scores for classification of cognitive impairment in paper III and IV were generated through quantile regression modelling with age, sex and level of education as predictor variables. Quantile modelling was applied to overcome issues of skewed data and issues of equal variance between groups of age, sex and education (Sherwood, Zhou, Weintraub, & Wang, 2016).

For our statistical evaluation of main effects, a p-value <0.05 was considered statistically significant. All analyses were conducted using IBM SPSS 22 or 25 with the exception that SAS 9.4 (SAS Institute, Cary, N.C.) was used for quantile regression modelling for creation of normative scores in Papers III and IV.

Table 4.
Statistical procedures applied in the four papers together with outcome variables and predictor variables

Papers	Aims	Main predictor variables (categorical levels)	Tests	Outcome	Other predictor variables	Main statistical procedures
Paper I	To assess influence of test setting factors on cognitive performance	Test administrator (21) Test order (2) Test version (3) Previous experience (2) Time of day (3)	Episodic memory, speed of processing spatial ability	Test scores on six tests and measuring three domains	Age (4) Sex (2) Education (2) Variation at participant level	Generalized linear mixed models
Paper II	To assess birth cohort effects on cognitive performance	Birth cohorts (3) Edcuation (3) Education (continuous)	Episodic memory, speed of processing, spatial ability, semantic memory, verbal fluency, attention and executive functioning, short-term and working memory and meta cognition	Test scores on twelve tests measuring six domains	-	ANOVAs and ANCOVAs
Paper III	To assess prevalence and incidence of MCI across age, sex and subtypes	Age (4) Sex (2)	Test for assesment of cognitive impairment: Episodic memory, speed of processing, verbal fluency, spatial ability	Prevalence/incidence rates of MCI	For the purpose of norm scoring: Age (3) Sex (2) Education (3)	Quantile regression (norm scoring) and Poisson regression
Paper IV	To assess stability of MCI diagnosis and assess factors associated with MCI reversion	*Set 1.* Demograhical factors: Age (4) Sex (2) Education (2) MMSE-scores (continuous) Marital status (2) *Set 2.* Psychological status: Depression (2) Motivation (2) Concentration (2) *Set 3.* Testing setting: Test order (2) Test version (3)Time of day(3) *Set 4.* MCI criteria: Amnestic (2) Severity (2) Domains (2) Bracket creep (2)	Test for assesment of cognitive impairment: Episodic memory, speed of processing, verbal fluency, spatial ability	Reversion rates of MCI	For the purpose of norm scoring: Age (3) Sex (2) Education (3)	Quantile regression (norm scoring) and logistic regression

Results

Paper I

The mixed linear models revealed a significant random effect corresponding to test administrators for all the cognitive tests (p<.01). The variation seen in test scores ascribed to the test administrator was between 1.4%-3.5% (see Table 5).

Table 5.
Variance estimates for different tests and variance components and percentage of variance corresponding to test administrator.

	Variance estimate	Standard error	Wald Z	*p*-Value	Percentage of Variance
Digit cancellation					
Test administrator	0.369	.129	2.86	.004	2.8%
Participant	9.07	.282	32.1	.001	
Residual	3.92	.118	33.2	.001	
Pattern comparison					
Test administrator	1.32	.437	2.82	.005	3.5%
Participant	22.3	.746	29.9	.001	
Residual	11.7	.355	32.9	.001	
Free recall					
Test administrator	0.065	.029	2.26	.024	1.4%
Participant	1.92	.099	19.4	.001	
Residual	2.59	.077	33.8	.001	
Recognition					
Test administrator	.021	.008	2.52	.012	2.6%
Participant	.314	.017	18.4	.001	
Residual	.473	.014	34.1	.001	
Speed of mental rotation					
Test administrator	1.64	.573	2.68	.007	1.5%
Participant	27.8	.851	32.7	.001	
Residual	71.6	.569	125.9	.001	
Accuracy of mental rotation					
Test administrator	.0006	.0002	2.49	.013*	1.9%
Participant	.008	.0007	11.6	.001	
Residual	.024	.0007	34.5	.001	

Further, different testing specific factors were associated with particular cognitive tests. There were differences between test versions for episodic memory, speed of processing and accuracy of mental rotation ($p<0.02$). Morning testing was beneficial for performance on memory tests ($F=14.4$, $df=46.9$, $p<0.001$), and being tested in the afternoon was beneficial for speed of mental rotation performance ($F=6.23$, $df=37.6$, $p<0.002$). Better performance scores were found when the speed of processing task (digit cancellation: $F=12.4$, $df=46.8$, $p<0.001$, pattern comparison: $F=87.8$, $df=23.0$, $p<0.001$) or speed of mental rotation ($F=42.8$, $df=35.0$, $p<0.001$) was administrated as the second speed task in the test battery. Participants that were tested at home performed worse on both speed of processing tasks and memory tasks compared to the participants that were tested at the research centre ($p<0.001$). Participants with previous experience of cognitive testing (a form of practice effect) performed better ($p<0.05$) on test with a speed component in comparison to those who had no prior experience.

Paper II

Birth cohort effects on test performance

ANOVAs evaluated test score difference on 12 cognitive tests among three birth cohorts for the 60- and 81-year-old groups. There were differences in cognitive test scores between birth cohorts in both the 60- and 81-year old groups. Differences were found for tests measuring speed of processing (digit cancellation and pattern comparison) and attention (TMT-A). Exclusively for the 60-year old groups, differences between cohorts in test scores of executive functioning (TMT-B), vocabulary, free recall, and recognition were observed. Means and medians were higher (indicating better performance) for cohort 2 and 3 in comparison to cohort 1 and post hoc tests (pairwise analysis) confirmed that this was the case where no significant differences between cohort 2 and 3 were observed. On further examination, using a contrast analysis, for 5 of 9 tests there was a non-linear trend where differences between cohort 1 and 2 appear to be unequal to differences between cohort 2 and 3. The results are presented in Table 6.

Table 6.
Analysis of Variance (ANOVAs) for differences in test scores between birth cohorts for 60- and 81-year-olds and for each cogntive test, pairwise comparisons with Hochberg GT2 corrections

				Cohort 1 vs. cohort 2	Cohort 2 vs. cohort 3	Cohort 1 vs. cohort 3	Contrast analysis for linearity
Cognitive test	*Welch (df1,df2)*	*MS*	Cohen's *d*	*Mean difference*	*Mean difference*	*Mean difference*	*t (df)*
Episodic memory							
Free recall							
60	3.78*(2,252)	17.5	0.19	-.443˚	-.002	-.445*˚	-.881 (124)
81	.594 (2,253)	2.71	0.14	-.26˚	.008	-.252	-.617 (239)
Recognition							
60	2.99 (2, 252)	19.7	0.18	-.616	.248	-.369	-1.55 (723)
81	.200 (2,235)	.172	0.06	-.079	.034	-.046	-.167 (415)
Speed of processing							
Digit cancellation							
60	6.66** (2,249)	104	0.25	-1.39**	.535	-.858˚	-2.22 (732)*
81	3.54* (2,243)	46.1	0.24	-1.27*	.647	-.622	-2.43 (424)*
Pattern comparison							
60	14.5*** (2,262)	124	0.38	-1.31***	.201	-1.11***	-2.3 (731)*
81	3.08* (2,249)	29.9	0.25	-.826	-.037	-.863*	-1.17 (422)
Attention *TMT-A* (in seconds)							
60	9.3*** (2,267)	128	0.32	1.24**	-.003	1.24***	1.44 (678)
81	4.61** (2,217)	181	0.31	2.02	.308	2.33**	1.22 (390)
Executive functioning *TMT-B* (in seconds)							
60	13.8*** (2,247)	760	0.43	3.87***˚	-.714	3.16***˚	3.18 (142)**
81	.833 (2,100)	207	0.18	.879	2.35	3.23	-.285 (203)
Semantic memory							
Vocabulary							
60	3.81* (2,262)	73.2	0.20	-1.35*	1.23	-.124	-2.67 (144)**
81	.839 (2,250)	24.4	0.13	.557	.260	.816	-.248 (426)
General knowledge test							
60	.499 (2,249)	1.23	0.06	-.058	.168	.110	-.633 (733)
81	.246 (2,233)	.552	0.06	-.009	.106	.097	-.345 (424)

Note. 60: participants were aged 60 at testing occasion. 81: participants were aged 81 at testing occasion. *$p < 0.05$ **$p < 0.01$, ***$p < 0.001$. Abbreviation: MS = Mean Square

Table 6 continue.

				Cohort 1 vs. cohort 2	Cohort 2 vs. cohort 3	Cohort 1 vs. cohort 3	Contrast analysis for linearity
Cognitive test	*Welch (df1,df2)*	*MS*	Cohen's *d*	*Mean difference*	*Mean difference*	*Mean difference*	*t (df)*
Word fluency							
Categories							
60	1.26 (2,260)	28.2	0.11	-.476	-.118	-.595	-.344 (729)
81	.031 (2,240)	.614	0.00	-.028	.120	.091	-.150 (424)
Letters							
60	1.63 (2,260)	36.0	0.13	-.684	.073	-.611	-.719 (729)
81	.341 (2,249)	.614	0.00	-.028	.120	.091	.809 (426)
Short-term memory							
Digit span forwards							
60	1.06 (2,249)	1.23	0.11	-.176	.134	-.042	-1.28 (728)
81	1.11 (2,247)	.933	0.14	.179	-.069	.109	1.19 (421)
Working memory							
Digit span backwards							
60	1.41 (2,264)	1.62	0.11	-.202	.160	-.042	-1.40 (725)
81	2.75 (2,247)	2.63	0.23	.027	.209	.236	-.852 (421)
Metacognition							
Confidence jugdement							
60	.173 (2,251)	.001	0.00	-.002	-.002	-.004	.037 (733)
81	1.37 (2,239)	.008	0.17	-.010	-.005	-.010	-.326 (424)

Note. 60: participants were aged 60 at testing occasion. 81: participants were aged 81 at testing occasion. $*p < 0.05$ $**p < 0.01$, $***p < 0.001$. Abbreviation: MS = Mean Square

Education

ANOVAs and ANCOVAs (for years of education) examined the role of education on birth cohort effects. Birth cohort effects remained significant for both tests measuring speed of processing, attention (TMT-A), and executive functioning (TMT-B) and test score differences were stagnant between cohort 1 and 2, and 1 and 3 respectively. The results were the same regardless of whether education was defined as a categorical variable or as a continuous variable.

Paper III

Prevalence

Overall MCI prevalence was 21.4% (95% confidence interval (CI) 20.1-22.8) for the lenient criterion and 6.6% (95% CI: 5.74-7.42) for the strict criterion. No significant sex-differences (MCI in men: 34.7%, MCI in women: 32.3%) (χ^2=0.324, *df* = 1, *p*= 0.299) were observed. Of the MCI subgroups, naMCIs (non-amnestic MCI single domain) had the highest prevalence (lenient: 11.5%, CI: 10.5-12.5, strict: 5.3%, CI: 0.134-0.521) and naMCIm (non-amnestic MCI multiple domain) had the lowest (lenient: 2.1%, CI: 1.67-2.61 strict: 0.327%, CI: 0.134-0.521). There were significant age-differences in prevalence estimates ($\chi^2 = 48.1$, df = 3, $p < 0.001$) See Figure 10 and 11 below for specific prevalence of MCI stratified by age and subtypes.

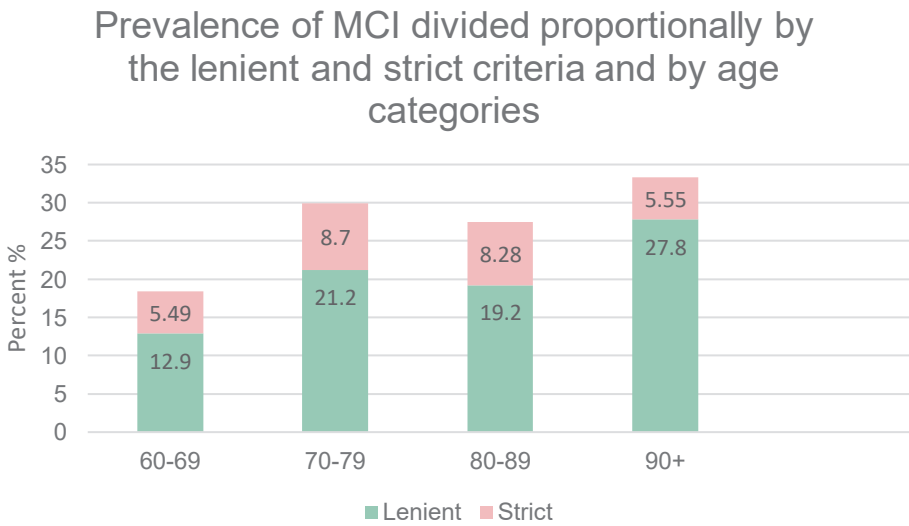

Figure 10.
Bar chart of age-stratified prevalence of MCI divided proportionally by the lenient and strict criteria.

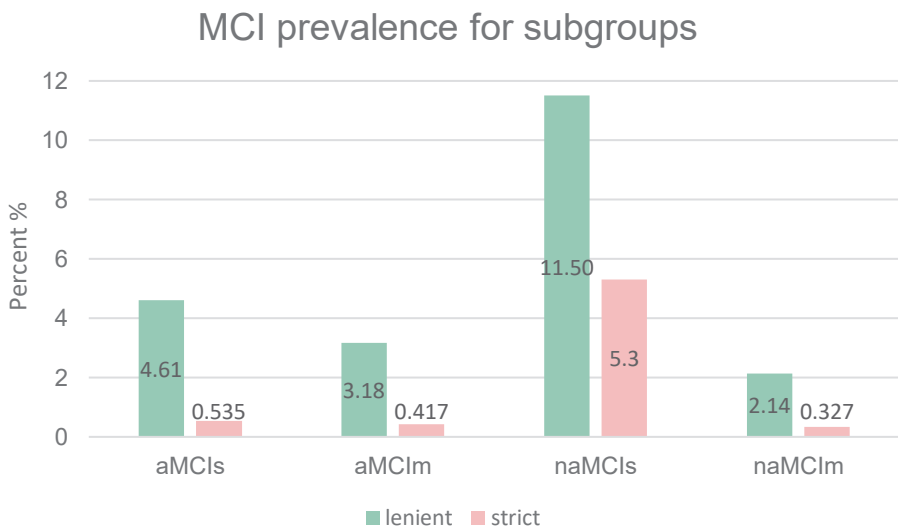

Figure 11.
Bar chart of MCI prevalence for subtypes

Incidence

The cumulative incidence was 13.9 % per 6 years (202 cases per 6 years) and the overall incidence rate of MCI was 22.6 (CI:19.6-25.9) per 1000 person-years for the lenient criterion. For the strict criterion the cumulative incidence rate was 5.15% per 6 years (85 cases per 6 years) and the overall incidence rate of MCI was 8.67 (95% CI: 7.0-10.7) per 1000 person-years. For MCI subtypes, naMCIs (non-amnestic MCI single domain) had the highest incident rate (lenient: 12.5, strict: 7.04), followed by aMCIs (amnestic MCI single domain (lenient: 6.70, strict: 0.918). See Figure 12 below for incidence stratified by age and amnestic and non-amnestic. See Table 7 for results from the Poisson regression.

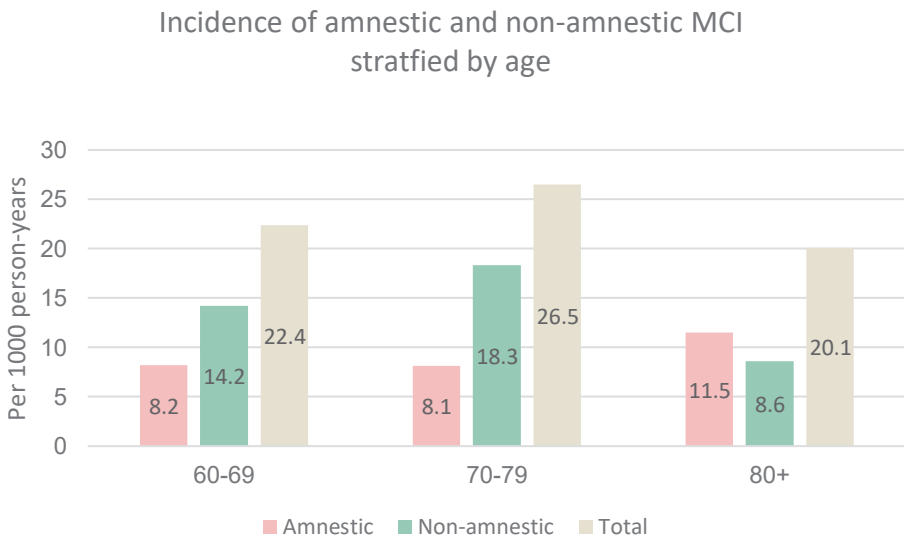

Figure 12.
Bar chart of Incidence of amnestic and non-amnestic MCI stratified by age

Table 7.
Overall MCI incidence rates for baseline samples for 6-year follow-up. Results from the Poisson regression analysis and overall MCI stratified by sex and age

	Lenient criterion						Strict criterion				
	n	Person -years	Incidence rate (95% CI)	*B* (95% CI)	*p*-value		*n*	Person-years	Incidence rate	*B* (95% CI)	*p*-value
60 age-group	155	6923.1	22.4 (19.1-26.2)	0[a]			60	7621.9	7.87 (6.11-10.1)	0[a]	.
70 age-group	26	982.8	26.5 (18.0-38.9)	0.167 (-0.248- 0.582)	.431		16	1052.2	15.2 (9.31-24.8)	0.658 (0.107- 1.21)	.019*
80+age-group	21	1047.0	20.1(13.1-30.8)	-0.11 (-0.566- 0.346)	.636		9	1129.2	7.97 (4.11-15.3)	.012 (-0.688-0.713)	.972
Women	103	4804.0	21.4 (17.7-26.0)	0[a]	.		47	4511.4	8.89 (6.68-11.8)	0[a]	.
Men	99	4148.8	23.9 (19.6-29.1)	0.107 (-0.169-0.383)	.447		38	5292.2	8.41 (6.12-11.6)	-0.053 (-0.48 - 0.375)	.808
Incidence rate	202	8952.9	22.6 (19.6-25.9)	.	.		85	9803.6	8.67 (7.0-10.7)	.	.
Incidence rate - no cognitive complaint	292	8952.9	32.6 (29.0-36.6)	.	.		113	9803.6	11.5 (9.6-13.9)	.	.

Note. 0a= reference group. *$p < 0.05$ **$p < 0.01$, ***$p < 0.001$

Paper IV

Trajectory and reversion of MCI

Over half of the participants (58%) with MCI reverted back to normal at 6-year follow-up. Of those with stable MCI, 43.5% remained within their subtype and 56.5% changed subtype. Of those who changed subtype, 16.7% went from single to multidomain whereas 20.5% went from multiple domain to single domain. 25.6% changed from amnestic to non-amnestic subtype and 37.2% changed from non-amnestic to amnestic. Of notice, of those in the latter conversion group, 62.1% further added an amnestic component to their MCI classification (i.e. single non-amnestic to amnestic multiple domain).

Our 12-year data revealed that of those participants with MCI at baseline and 12-year follow-up, 48.5% transitioned back to normal cognitive functioning at 6-year follow-up (unstable trajectory), whereas the remainders (51.5%) had a stable MCI trajectory throughout the 12-year examination period.

Factors tied to reversion

Five separate logistic regression models were run, including sets of factors associated with either demographical factors (model 1), psychological status (model 2) testing session factors (model 3) or MCI criteria (model 4). The fifth and final logistic regression model included the significant predictors found in the aforementioned analyses. The four first models found that lower age, better global functioning, good concentration and single domain MCI were associated with reversion. The same factors were also significant predictors of reversion in the fifth model. See Table 8 for results from the fifth and final logistic analysis.

Table 8.
Results from the final logistic regression analysis with reversion as outcome

		Model 5		
Factors		OR exp(b)	95% CI	*p*-value
Demographical factors				
sex	male	.820	.513-1.31	.409
	female	0[a]		
Age groups	60s	2.19	1.08-4.43	.030*
	70s	3.11	1.27-7.62	.013**
	80+	0[a]		
Global cognitive functioning	MMSE	1.15	1.03-1.29	.018*
Psychological status				
Concentration	Good	2.49	.992-6.23	.052*
	Bad	0[a]		
Single or multiple domain	Single	4.98	2.22-11.1	.000***

Note. $p<0.001$.MMSE= Mini mental state examination. 0a= reference group, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

Discussion

Summary of main findings

The main finding of paper I was the quantification of variation in participants' test scores due to the factor corresponding to the test administrator. This is interpreted as test administrator influence, and was seen for tests measuring episodic memory, speed of processing and spatial ability. Secondary findings included enhanced memory scores for morning examinations, differences in test scores attributable to test versions and order of test administration, individuals tested at home performed poorer than those tested at the clinic, and finally, previous experience of cognitive testing (practice effects) was beneficial for performance on tests with a speed component. The main findings of paper II were that later born cohorts (cohort 2 and 3) outperformed the earlier born cohort (cohort 1) for tests measuring speed of processing and attention. Level of education could not fully explain cohort related gains. Papers I and II provide evidence for instrumentation and cohort effects on participants' test scores, thus concluding that threats to internal validity are tangible concerns in ageing studies that examine cognition.

The papers examining mild cognitive impairment found that ≈20% of the sample aged 60 and over had some form of MCI and ≈7% had MCI according to the strict criterion for cognitive impairment (two impaired test scores). Incidence for overall MCI was 22.6 and 8.67 per 1000 person-years, for the lenient and strict MCI criterion respectively, which is comparable to what is reported in other MCI studies. No sex differences were found for occurrence of MCI. Prevalence of MCI increased with age. However, age-effects on incidence rates were inconclusive. Over half of the participants with MCI at the initial examination were classified as cognitively intact (reverters) at 6-year follow-up. A little over half of the non-reverters (e.i. stable MCI) changed their subtype classification. The reverters had lower age, better global functioning, better concentration and single domain subtype (vs. multiple domain) at baseline assessment compared to those with stable MCI. Interpretations of the results from these two latter papers are that prevalence and incidence estimates are highly dependent on which operationalisation of cognitive impairment is applied and that MCI is a fairly unstable diagnostic entity with a substantial number of individuals with MCI reverting back to normal or changing subtype.

Interpretation of results

Test administrator influence

The focus of paper I was to investigate test administrator influence on cognitive testing. There was variation seen in participants' test scores attributable to the test administrator, although the proportion of the total variance was interpreted as fairly small. The small effect size could be clarified by the rigid routines that the test administrators in GÅS follow. High quality training, extensive test manuals and meetings regarding methodological discrepancies in testing are examples of precautions taken to minimise variability in test administration.

The enquiry as to whether cognitive testing is more vulnerable to test administrator influence when using samples of older adults than when using younger samples has been considered previously (Lezak et al., 2004; Sattler & Theye, 1967) but, to the author's knowledge, not been tested. The basis for this notion stems from the understanding that cognitive ageing (i.e. natural decline in cognition), as well as the common difficulties with hearing and eyesight among older adults, may interfere with the elder's comprehension of the test instruction, consequently leading to departures from standardised procedures. Supplementary instructions may vary between test administrators, either causing or amplifying systematic test administrator influences. Therefore, it is suggested that large studies, especially those testing older adults should routinely check for test administrator influences.

Characteristics of the administrator have been repeatedly shown in the literature to bear partial responsibility for how well the test-taker performs on a cognitive test (e.g Chapman et al., 2018). Lack of experience and sex of the test administrator are examples of such characteristics. Chapman and colleagues (2018) investigated the role of the administrator's sex in a plethora of contexts and recurrently found in the literature that performance, be it physical or mental, was better when the performer was paired with the opposite sex experimenter. There were only two male test administrators in GÅS, making it difficult to draw conclusions regarding whether the sex of the test administrator caused the observed variance. The other characteristic of interest pertains to the amount of testing experience the administrator has. Inexperienced test administrators make more mistakes when reciting test instructions and when correcting tests, and they may also come across as nervous compared to experienced ones, which can make the participant nervous causing them to perform poorly. Previous research has provided evidence indicating that lack of experience leads to discrepancies in test routines between test administrators (Hoyt & Kerns, 1999). When separate analyses on our data involving only experienced test administrators were performed, significant variations in test scores attributable to the administrators remained (data not shown). Therefore, it is

argued that the experience of the test administrator cannot, at least not fully, explain our results.

Future perspectives

As the population is getting older, more cognitive testing will be carried out on this specific population group. Future research could investigate to what extent older adults are more prone to evoke deviation from standardised procedures, because of difficulties in comprehension of test instructions, in comparison to younger adults.

A further research objective could involve the digitalisation of tests, which eliminates specific test administrator influence (Germine, Reinecke, & Chaytor, 2019). Digital test-taking could also reduce other potential variation caused by fatigue or nervousness through a method called *burst testing*. This entails seven days of repeated testing through digital tests administrated via smart phones. An average score is then assembled into one test score to gain a more accurate overview of the level of cognitive ability (Sliwinski et al., 2018). High re-test reliability has been found with this technique and can overcome problems such as misclassification of cognitive impairment.

Birth cohort effects

The results from paper II add to the existing evidence that cohort effects are prominent in samples of older adults. The results show that speed of processing and attention (measured with TMT-A, occasionally used as a measure of processing speed) were the cognitive abilities most subjectable to birth cohort effects. Previous research is consistent with these results. For example, Thorvaldsson, Karlsson, Skoog, Skoog, and Johansson (2017), using a Swedish sample aged 70-79, found cohort effects on test scores measuring perceptual-motor speed. Moreover, Dickinson and Hiscock (2011) found substantial cohort related gains for normative test scores on TMT for those aged 30-70. In addition, when examining data from the Seattle longitudinal study, Salthouse (2015) observed gains for perceptual speed. Nonetheless, Finkel et al. (2007) failed to find cohort effects on speed of processing measures when examining Swedish twins aged 50 and over. Differences in test-taking patterns may explain the observed cohort-related gains (Brand, 1996; Must & Must, 2013), where later born cohorts have been taught to emphasise speed over accuracy, whereas earlier born cohorts have a tendency to do the opposite. The magnitude of observed gains in our data is comparable to previous studies with modest effect sizes (Skirbekk et al., 2013).

No cohort differences were detected for measures of short-term memory and working memory, coherent with previous findings (e.g. Dickinson & Hiscock, 2011; Gignac, 2015). Additionally, no cohort differences were found for meta cognition,

which is a novel result as this relationship has not previously been investigated. The instable cohort effects between analyses for episodic memory makes it problematic to draw stern conclusions regarding whether differences between cohorts were arbitrary or true cohort related gains.

Birth cohort effects remained significant after controlling for length and level of education. Education has been favoured as a cause of birth cohort effects in much research (Pietschnig & Voracek, 2015; Rönnlund & Nilsson, 2008) which makes these results somewhat surprising. However, there are previous reports using Swedish samples where education could not fully account for the observed effects on multiple cognitive measures (Karlsson et al., 2015). We did not test for other explanatory factors which makes it difficult to determine the underlying causes of effects. Rönnlund and Nilsson (2008), using a Swedish sample, reported that over half of the observed cohort effects were explained by level of education. Sibling size and nutrition were secondary explanatory factors and these factors could be applicable explanations to our results as some of our cohort samples were overlapping in birth years. The cohort effects observed in this study highlight the role of environmental factors on cognition and that researchers need to consider these when establishing life-span trajectories.

Future perspectives

Cohort effects have also been proven present in prevalence and incidence estimates of dementia (Dodge et al., 2017; Dodge, Zhu, Lee, Chang, & Ganguli, 2014; Fratiglioni et al., 2017; Kosteniuk et al., 2016; Seblova et al., 2018) as well as in the rate of cognitive decline (Gerstorf, Ram, Hoppmann, Willis, & Schaie, 2011; Skirbekk et al., 2013; Thorvaldsson et al., 2017). In light of these findings, cohort effects on estimates of MCI and whether rate of decline from MCI to dementia is differs between cohorts could be a future direction for research.

Additionally, education is typically included in statistical models, as a mediating variable, to eliminate unwanted cohort effects in cognitive cross-sectional data. Though, the results from paper II, and other studies, indicate that this procedure is not always sufficient (Dodge et al., 2014; Karlsson et al., 2015). Thus, it is imperative for researchers to continue to search for potential factors explaining cohort related gains, in order to fully utilise cross-sectional data in ageing studies.

Mild Cognitive Impairment

Prevalence and incidence

Our prevalence estimates are within the previously reported 3.2-42% (Petersen et al., 2018; Ward et al., 2012) and our lenient prevalence estimate is similar with average estimate of 18.9% reported by Petersen et al. (2014) in a sample of adults

aged 50 and over. Incidence rates for overall MCI were also within the previously reported 3.2-104.6 per 1000 person-years (Luck et al., 2010; Manly et al., 2008; Roberts et al., 2014; Solfrizzi et al., 2004). When combining single and multiple domain amnestic MCI, incidence rate for aMCI was 8.6 per 1000 person-years, which is below the estimate of 14.2 reported another study using GÅS-participants (Elmståhl & Widerström, 2014). Differences in diagnostic criteria may explain deviation in results. There were more non-amnestic measures than amnestic in our study, which serves as an explanation of naMCI producing the highest estimates. In addition, single domain MCI tend to generate the highest estimates (Brodaty et al., 2013; Manly et al., 2008; Palmer, Bäckman, Winblad, & Fratiglioni, 2008), consistent with the results in our study.

The secondary aim in paper III was to report dissimilarities in estimates of MCI occurrence due to different criteria, and in this case, there were two criteria, a strict (≥1 impaired test score) and a lenient (≥2 impaired test scores) for cognitive impairment. A difference of 14.8 percentage points in prevalence was found. Substantial variations in both prevalent and incident MCI for 2 vs. 1 test to define impairment have been described previously (Brodaty et al., 2017; Brodaty et al., 2013; Jak et al., 2016). For example, when comparing the Jak/Bondi criteria (with a strict two tests per cognitive domain and a cut-off of ≥1 SD below norms) with the Mayo Clinic/Winblad criteria (with a less strict one test per cognitive domain, and a cut-off of ≥1.5 SD below norm scores) a discrepancy of 10 percentage points of prevalent MCI was found. With this said, few studies describe criteria discrepancies using a 1 vs. 2 test classification. The number of impaired test scores could be a proxy for the severity of impairment and thus predict risk of future dementia or worse cognitive status, which was the case in two studies that used a strict criterion of two impaired scores (Loewenstein et al., 2009; Manly et al., 2008). Contradictory to this, Jak et al. (2016) failed to find a difference in the prediction of future dementia between the Jak/Bondi criterion or the Petersen/Winblad criterion. Moreover, paper IV did not find an association between stable MCI and the strict criterion.

Age and sex

Age patterns for prevalence and incidence estimates were inconclusive. Consistent with previous studies, prevalence of MCI increased with age, with the lowest and highest prevalence found for the 60 and 90+ groups respectively. Remarkably, no steady increase per ascending age group was observed, which is contradictory to the conclusion from Petersen et al. (2018) and Gillis et al. (2019) meta-analyses and several other studies (Caracciolo et al., 2008; Luck et al., 2010; Manly et al., 2008). Nevertheless, there have been reports stating that MCI does not necessarily increase with age and it can depend on the diagnostic criteria applied. For instance, Trittschuh et al. (2011), noted that when applying a less strict criterion (one impaired

test score) prevalence estimates were similar across all age group. What is more, prevalent amnestic MCI was reported to be stable across age groups in Ward et al. (2012). Thus, the age-MCI relationship may be far more complex than suggested. Our incidence data for different age groups are consistent with research reporting no successive increase with age (Brodaty et al., 2013; Busse et al., 2006; Larrieu et al., 2002; Ravaglia et al., 2008). A mixture of selection bias, the application of age corrected normative scores and the age-bracket hypothesis (Aerts et al., 2017) may explain why we failed to find age dependent incidence rates. Furthermore, our results did not support the notion that occurrence of MCI is higher for women. The lack of observed sex differences for MCI estimates is consistent with the results of a meta-analysis conducted by Au et al. (2017) and, in part, the latest report from Petersen et al. (2018) of combined amnestic and non-amnestic type using population-based samples and the MCI Mayo Clinic/Winblad criteria. Conversely, previous reports on incidence using GÅS data have shown higher estimates for men (Elmståhl & Widerström, 2014).

Reversion rates

The 6-year reversion rate in paper IV was higher than the average of 29%, for both amnestic and non-amnestic subtypes, reported in a population based meta-analysis conducted by Canevelli et al. (2016), but within the reported range of 4%-58%. There are several potential reasons as to why our rate was among the highest reported so far. Firstly, attrition, where the frailest (with MCI) left the study prior follow-up. In favour of this idea, baseline MCI participants who dropped out had lower MMSE scores and were older than those who stayed on. Notably, the participant rate for 6-year follow-up from Wave 1 and Wave 2 was 83.5% and 80.6%, respectively, which could be considered fairly high. Secondly, our sample was drawn from a population-based study. Higher reversion rates are found in population-based samples compared to clinical ones. Thirdly, the non-reverter group comprised exclusively of persons with stable MCI, and did not include dementia progressors. Occasionally, reversion studies group participants with mixed diagnoses (i.e. stable MCI with dementia progressors) together (Pandya et al., 2016). This complicates interpretation of results and may bring reversion rates down. Fourthly, a large proportion of our sample were younger (72.5% were 60-69) and older individuals (70+) are more likely to have mixed aetiology (e.g. mixed AD/Vascular features), which in general leads to faster progression compared to a single aetiology/pathology (Petersen et al., 2014). Additionally, younger participants from the community are also more likely to have treatable causes than very early dementia, (e.g. depression or side effects from medication). Finally, a large group of participants were excluded due to incomplete cognitive datasets (making it challenging to classify them as MCI/NCI). Participants with unfinished cognitive testing probably have some form of cognitive restraint, and by excluding these participants reversion rates are exaggerated. Remarkably, we had a relatively

long follow-up in comparison to other studies that report reversion rates. Longer follow-ups tend to generate modest reversion rates, due to the unstable MCI groups becoming stable or progressing to dementia when more time has passed (Lopez et al., 2012; Malek-Ahmadi, 2016; Roberts et al., 2014).

We found that it was fairly common to change subtype, with 56.5% of our MCI stable sample classified with another subtype at 6-year follow-up. This result is comparable with a previous Korean study (Han et al., 2012), which reported that 48% changed subtype. Subtypes are advantageous when figuring out the underlying cause of MCI diagnosis. By assessing the aetiology, the observer can predict which future dementia outcome is most likely. Excessive rates of conversion from one cognitive domain to another undermines the ultimate purpose of subtypes. Thus, further research on factors associated with different subtype progression and conversion is necessary.

Factors associated with reversion

Baseline predictors for reversion such as lower age, better global functioning, and single domain subtype (vs. multiple domain) have all previously been associated with reversion (Pandya et al., 2016). That these factors are all associated with cognitive health is not surprising as their opposites are well-known risk factors for conversion to dementia (e.g. Winblad et al., 2016). Difficulties with concentration has previously been associated with incident MCI (Lobo et al., 2008) and our results similarly show that good concentration could predict reversion. This particular association is a novel finding, and concentration is now an added predictor in the research of MCI reversion. Having a less severe (lenient criterion) form of MCI was not associated with reversion. This is surprising as it was theorised that if an individual had two impaired scores within the same domain then this was a marker of true cognitive impairment (affiliated with stable MCI group). Having one impaired score on the other hand could reflect a haphazard bad test score, leading to an incorrect MCI diagnosis which was detected at follow-up (affiliated with reverter group). Similar to our results, Jak et al. (2016) found no differences (15.95% vs. 12.78% $p<0.009$) in reversion rates when using one impaired test scores or two test scores within the same domain for cognitive impairment. Why less severe MCI could not predict reversion remains unanswered.

Future perspectives

In order to produce accurate prevalence and incidence estimates of mild cognitive impairment, consensus on a detailed operationalisation of the definition should be reached and put forward together with important biological markers of dementia. In addition, it is imperative to continue with the search for predictive factors for reversion, as this type of research has only begun to emerged. Moreover, MCI is a risk factor for dementia, even for those with an unstable trajectory (e.g. MCI-NCI-

dementia/MCI-NCI-MCI-dementia). By studying the characteristics of these unstable individuals in relation to individuals which never retransition back to MCI (MCI-NCI-NCI) could further help unravel risk factors tied to different MCI trajectories and ultimately predict which individuals with MCI are at the most risk to progress to dementia.

Methodological considerations

All four papers included investigation of how methodological implementations could influence cognitive outcome. In paper I and II, internal validity issues on participants' cognitive tests were examined. In paper III, the operationalisation of cognitive impairment was examined in relation to occurrence of MCI. In paper IV, test setting factors such as test order/version and time of day, and definition of MCI was examined in relation to reversion rates. In this section other methodological issues are discussed, that are of relevancy in epidemiological investigation and in the current papers. The issues are presented in the following order: selection bias, misclassification, generalisability, practice effects, using norm scoring to assess cognitive impairment, issues in comparison of research results and algorithm vs. consensus panel in MCI classification.

Selection bias, misclassification, generalisability and practice effects

Selection bias

Selection bias is described as a systematic error that occurs in a study due to the practices applied to select participants and from reasons that impact participation in the study (Rothman & Hatch, 2002). Selection bias should be considered as a potential issue in both the context of who agreed to participate in the beginning of the study, and who remain in the study (concerning concepts such as attrition and survival bias). Selection bias and attrition threatens the sample's representability of the general population, because the most unfit participants usually do not participate. For instance, the Maastricht Aging Study found lower baseline neuropsychological test scores among those who died or dropped out before 3-year follow-up in comparison to participants who stayed in the study (Van Beijsterveldt et al., 2002).

In order to reduce selection bias in the GÅS-study, participants that were too unfit to come to the research clinic were offered home visits. In paper I, results indicated that those who were tested in the clinic performed better than those tested at home. Although, a strict exclusion criterion (e.g. eliminating dementia cases) was applied in attempt to even out cognitive differences between the home-visit group, and the

research clinic group, a selection of the unhealthiest participants was probably still present in the home visit group, amplifying outcome.

In papers III and IV, incidence rates and reversion rates, as well as the independent variables' (e.g. age) weight on rates may have been distorted due attrition (i.e. a selection of the healthiest are examined at follow-up). This could most likely lead to underestimations of MCI, higher reversion rates and subdued aging effects on MCI incidence. In addition, prevalence estimates as well as age effects on prevalence could be underestimated because of selection of participants entering the study. Notably, compared to many other studies, our participant rates for both baseline samples and returning participants were fairly high, although generally participants rates in epidemiological studies are declining (Galea & Tracy, 2007). Hence, selection bias may only have a small impact on results presented in these papers.

Contrarily, fairly strict exclusion criteria were applied in all the papers, which is also considered a form of selection bias. As a consequence, in paper I and II, test administrator effects and cohort effects were most likely restrained, in paper III and IV, there was most likely an underestimation of MCI occurrence and reversion.

Misclassification

There may be issues of misclassification of both outcome (e.g. MCI estimates) and exposure variables (e.g. depression or concentration) in papers III and IV which could lead to distortion of the observed results, sometimes exaggerating and sometimes underestimating results. Differential misclassification is perhaps an issue in paper IV where the classification of the exposure variable depression is dependent on the outcome variable MCI. Depression often coincides with MCI and sometimes individuals with depression may be mistaken for having MCI. That exposure and outcome variables are related with each other can cause dependency in misclassification. To what extent the misclassification of exposure and outcome variables has influenced results is difficult to infer. Moreover, factors tied to misclassification of baseline MCI (e.g. temporary depression or using a lenient criteria) were investigated in paper IV. However, these factors were not associated with reversion. Nevertheless, there may still have been single cases were these factors did contribute to misclassification of baseline MCI. Non-differential misclassification, i.e. random misclassification is equal in different groups (e.g. age or education groups), could be an issue in papers III and IV. To what extent non-differential misclassification is present is unclear. Misclassification of MCI could be reduced by using multiple evaluations close in time instead of using a single point estimate (Gillis et al., 2019) or by using a consensus approach with multiple evaluators.

There might have been a slight underestimation of dementia in the GÅS population samples. For instance, prevalence of dementia worldwide is 5-7% (Prince et al., 2013) and in the other SNAC-studies it has been reported to be about 7% (Fratiglioni et al., 2017). Prevalence of dementia in paper III was 4%. If dementia diagnoses were missed in GÅS, this could result in overestimations of MCI, due to dementia cases being misclassified as MCI cases. However, all participants with impaired ADL were removed from our samples, and as dementia diagnosis is contingent on impairments of ADL, classification of MCI cases due to true dementia seems unlikely. Moreover, there was an oversampling of younger participants (ages 60-69), which could explain our low prevalence of dementia.

Generalisability

There are strengths and limitations in the four papers, which generates mixed conclusions regarding the generalisability of results to other study populations. Random sampling occurred in both rural and urban areas with fairly high participant rates (average participant rate of 72% for both baseline waves and follow-up samples), strengthening generalisability to other Swedish population samples. Participant rate was however reduced in the older age groups (e.g. 53% in 80-89 and 43% in 90+ for Wave 1), making generalisation of results for the oldest old more problematic, which is unfortunately a common drawback in ageing studies (Davies et al., 2010). Moreover, in Skåne and especially in Malmö, there is a fairly high number of foreign-born participants (17% in Skåne and 31% in Malmö, Statistiska Centralbyrån, 2013). In the GÅS-project about 10% of the baseline sample (Wave 1) is born outside of Sweden, challenging the representativeness of ethnic diversity in our sample. The non-participation of this group could probably be explained by language difficulties and other cultural differences influencing study participation. Moreover, during the studies' exclusion processes, foreign born participants with language difficulties were also excluded further reducing the number of participants born outside of Sweden in our sample. When cognitive data from SNAC-Blekinge and SNAC-GÅS were compared no overall differences in participants' test scores were observed, strengthening GÅS-data generalisability nationally. It seems problematic to apply the papers' results globally, as there are major distinctions between countries regarding factors that promote cognitive health, such as income, education and health care systems. Therefore, it is argued that our results are generalisable to samples of adults 60+ in other high-income countries with lower proportions of foreign-born individuals in north Europe.

Practice effects

In longitudinal data, retest effects are of particular interest, where participants improve their performance on various cognitive tests for each testing occasion. Better performance could then hide true change in cognitive ability between test occasions. Paper I found previous experience of the testing situation to positively

influence test results on two speed tasks, but not influence the other tests such as memory or accuracy of spatial ability. Due to this, the retest factor was not considered a factor when utilising longitudinal data in papers III or IV. There were supplementary cognitive tests (other than those measuring speed) applied to define cognitive impairment, but it remains unknown whether rehearsal of these tests was beneficial for performance. Practice effects could potentially lower incidence rates and push reversion rates up, masking true deterioration of cognition.

Even in cross-sectional designs, practice effects can bias results. Paper I found that when performing the second speed of processing task, participants performed better than when receiving the first task. To neutralise the problem (Ferrer & Ghisletta, 2011), the order of test administration was spread out equally across participants. Hence task familiarity issues should not bias our results in paper II where a cross-sectional design was used. As for papers III and IV the returning participants received the same order of test administration, so if an improvement was present in the first test occasion, it should be present in the next on the same test for the individual participant, counterbalancing the potential influence of task familiarity issues. Moreover, there were no differences in the test order between reverters and non-reverters. Dodge et al. (2017) reported cohort differences in practice effects on verbal memory, where later-born cohorts demonstrated a superior capability to learn from prior testing. This adds an additional dimension to the study of cohort effects on cognitive performance using longitudinal data, where previous experience of cognitive testing for different cohorts should be considered.

Other methodological considerations

Using norm scoring to assess cognitive impairment

Norm data used in papers III and IV may have been confounded with cohort effects, as we used baseline data collected in 2001-2004 (early born cohorts) to create normative scores and then applied them to data collected in 2006-2012 and 2012-2017 (later born cohorts). This might consequently lead to underdiagnoses of MCI in the later born cohort due to the usage of potentially out-dated norms. There are however a few reasons as to why not to raise concern. Firstly, results from paper II revealed cohort effects for speed of processing tasks only, leaving a very small impact on MCI diagnosis. Secondly, we applied norm scores corrected for education, potentially reducing cohort effects. Thirdly, no differences in MCI prevalence between cohorts were observed.

Deciding which inclusion criteria to apply for norm samples can be tricky as the healthier the sample, i.e. the stricter the inclusion criteria, the higher the proportion that will be classified as cognitively impaired. Individuals with MCI were not excluded from the norm-samples in papers III and IV, probably leading to fewer

classified as cognitively impaired, compared to the case where MCI individuals were excluded from norm sample. It is nonetheless debatable whether individuals with MCI should be part of a normal healthy ageing population.

Issues in comparison of research results

Comparison of research results from different studies is an essential part of research and a way to try to validate your own results. There are many pitfalls when making comparisons, and methodological differences can explain a large amount of result discrepancies. Therefore, caution should be taken when drawing parallels.

Trahan et al. (2014) highlight this problem in the context of cohort effects, where major differences involve differences in length of time spans between cohorts, the nature of the samples (e.g. including longitudinal data or not), single vs. aggregated cognitive measures and, single vs. aggregated birth cohort samples. As long as there is a fluctuation in study methods, it will be challenging to directly compare results and comprehend what cognitive domains and which cohorts are temporally improving.

As for prevalence, incidence, reversion rates and risk factors for MCI, Petersen and colleagues (2014) describe several dissimilarities between studies. These being: the source and age of subjects, implementation of definition, retrospective vs. prospective data collection, algorithmic vs. clinical application of MCI classification (discussed in the next section), blindness of previous diagnosis/cognitive status and length of follow-up. Meta-analytic studies of prevalence and incidence of MCI try to reduce heterogeneity in results through selecting studies that have similar characteristics, e.g. they use the same definition and criteria for MCI, same age categories or have similar follow-up lengths. Despite applying restrictive inclusion criteria, these meta-analytic studies still report considerable heterogeneity in the occurrence estimates of MCI (Au et al., 2017; Gillis et al., 2019; Ward et al., 2012). The observed inconsistencies in MCI outcomes further promotes the need to come to a consensus regarding the definition and operationalisation of MCI. More studies applying multiple MCI definitions in relation to predictive power of progression to dementia may untangle which definition is most suitable as a prodromal stage of dementia.

Algorithm vs. consensus panel in MCI classification

Algorithm vs. consensus panel in MCI classification is one main methodological variability between studies (Petersen et al., 2014). Algorithmic classification is typically applied retrospectively on population-based data, whereas the consensus approach is applied on clinic-based subjects and it closely represents the actual practice of diagnosis. Algorithm has the advantage of being reproducible and objective in classifying participants, and its emphasis lies on the neuropsychological test scores. Consensus has the advantage of incorporating all medical aspects of the

individual, making this approach more accurate in the identification of true pre-stages of dementia (Malek-Ahmadi, 2016; Shadish, Cook, & Campbell, 2002). This approach nevertheless suffers from subjectivity, with inconsistencies in clinical judgements, much like test administrator influence. The background of the specialist or the amounted experience of practitioner etc., are aspects that leave an imprint on the inferences of the participant or patients' level of cognitive impairment. Petersen et al. (2014) suggest a type of algorithmic approach taking important medical aspects into consideration in a standardised manner, with special emphasis on medical history and temporal changes is the way forward in the diagnostic definition of MCI. However, the research literature suggests that neuropsychological tests alone can accurately predict future dementia when using an algorithmic approach. Therefore, researchers as well as clinicians should not neglect the importance and the utility of neuropsychological testing.

Clinical implications

The results from paper I have highlighted the importance of maintaining similar test settings. Applying these results in a clinical setting could entail that patients should be examined by the same clinician during cognitive investigation, e.g. the evaluation of dementia, in order to reach a fair conclusion regarding cognitive level, especially if cognitive testing is involved. One way of reducing the inevitability of human error in testing is by using computer-based testing. Some clinicians and clinics have already started to implement digitalisation of cognitive testing (Germine et al., 2019), and as our world is becoming more digitally oriented, so will health care. With regard to cohort effects in clinical settings, environmental factors become extremely important when using normative cognitive scores to establish cognitive status for the purpose of dementia diagnosis as the normative scores are susceptible to secular changes.

Despite the conclusions from the papers in this thesis, namely that MCI is a heterogeneous condition, with a bidirectional trajectory, I nevertheless argue that MCI is a useful concept for researchers and clinicians. Many studies using clinical samples have applied mild cognitive impairment successfully, with low reversion rates with high sensitivity and specificity for dementia (e.g. Belleville et al., 2017; Belleville, Gauthier, Lepage, Kergoat, & Gilbert, 2014). What is more, MCI was first designed as a clinical tool which may explain the number of unsuccessful attempts to define preclinical stages of dementia in the population-based samples. The results from papers III and IV can contribute towards the awareness among clinicians that cognitive impairment is fairly common in the general population and that it does not always necessarily lead to further cognitive deterioration. This awareness can provide a more cautious perspective, leading to the prevention of

overdiagnosis of cognitive disorders, subsequently avoiding the stigma that comes with being labelled as cognitively impaired (Garand et al., 2009). Awareness of factors associated with reversion are beneficial for clinicians or home-carers in order to set appropriate expectations pertaining future cognitive functioning, and consequently leading to better planning for future care (e.g. level of monitoring or determining the need for medications). In addition, the findings on reversion factors can aid participant selection for intervention studies that intend to prevent or delay onset of dementia.

Conclusions

Knowing what internal validity threats potentially influence cognitive outcome is an essential aspect of the practice of good research. Often, methodological practicalities are implemented to reduce these threats, such as changing test versions or order of test administration. These implementation strategies are sufficient to satisfy the researcher's concern of validity issues and therefore further investigation of potential methodological threats are sometimes neglected. Undeniably, certain issues are more pronounced and investigated in the literature than others, e.g. retest effects are well-documented, whereas descriptions of test administrator effects in ageing research are scarce. The results of paper I and II clearly demonstrate that instrumentation, practice effects and cohort effects are present in a large population study of cognitive ageing. Therefore, I argue that bias in results linked to internal validity should be routinely investigated in ageing studies.

There appears to be an ongoing development of two opposing stances within MCI research. One stance recognises MCI to be an unstable concept, with a bidirectional outcome and excessive questions regarding the implementation of the MCI definition. This leads to diverse reports regarding prevalence, incidence, stability, trajectories, risk factors, protective factors and factors tied to reversion. The other stance argues that MCI is a helpful diagnostic entity which can accurately predict worsening of cognitive status. Our results provide support for both stances. For the first account, our papers reported high diversity in prevalence and incidence estimates and high reversion rates as well as an unstable MCI trajectory where more than half of the MCI participants changed subtype at follow-up. Our data also provides support for the second notion as the results in paper IV implied that those who had MCI at baseline were more likely to have MCI at 12-year follow-up compared to those who never had MCI. This fact makes it difficult to fully advocate that MCI is a heterogeneous concept with a bidirectional outcome lacking relevance in predicting future cognitive status.

In conclusion, correctly assessing cognition in older adults comes with many methodological considerations, both clinically and in research. The results presented in this thesis have shown that every aspect in the assessment can play a major role in the outcome of cognition, be it the decision of what is considered cognitive impairment, or figuring out when cognitive ageing onsets. All stages of assessment are relevant, ranging from who collects the cognitive data, to taking the test-taker's birth date into consideration, to determining the best criteria for assessing impairment so as to accurately predict future dementia. My hope is that the research presented in this thesis can encourage researchers to reflect over methodological choices and how they can interfere with interpretation of cognitive outcome. This research also has ambitions to contribute towards finding an optimal way to define a prodromal stage of dementia.

Sammanfattning på svenska

Bakgrund

Med en snabbt växande äldre populationsgrupp kommer även andelen med kognitiva störningar, som demens, att växa. Det är således nödvändigt med forskning på kognition och åldrande. Denna avhandling, genom fyra delarbeten, redogör för testledareffekter vid kognitionsmätningar och hur yngre födelsekohorter (generationer) presterar bättre än äldre på kognitionstester. Den kommer även förklara prevalens (förekomsten), incidens (uppkomsten) och tillbakagången (förbättringen) av lindrig kognitiv störning (som anses vara ett förstadium till demens).

Arbete I och II testledareffekter och kohorteffekter

Kognitiva tester, vilka används för att mäta kognitiva störningar, är känsliga för mätproblem och kan orsaka att det kognitiva utfallet (t ex. kognitivt testresultat) inte överensstämmer med verkligheten (t ex. deltagarens kognitiva nivå). *Testledareffekter* och *kohorteffekter* är exempel på faktorer som kan orsaka mätproblem vid kognitiv testning.

Testledareffekter innebär att personen som administrerar testerna kan påverka hur deltagaren eller patienten presterar på ett kognitivt test (med positivt eller negativt utfall). I stora befolkningsstudier används ofta multipla testledare, vilket kraftigt ökar risken för oönskad variation i deltagarnas testresultat.

Kohorteffekter eller födelsekohorteffekter innebär generellt en påverkan på ett utfall och i detta specifika fall testresultatet på ett kognitionsinstrument utifrån när provtagaren är född. Forskningen har visat att yngre födelsekohorter presterar bättre på kognitionstester, i genomsnitt, än äldre födelsekohorter. Oftast förklaras denna skillnad med att yngre födelsekohorter har högre utbildningsnivå jämfört med äldre kohorter. Det är därför viktigt att ta hänsyn till när deltagaren är född vid åldersjämförelser i kognitionsnivå.

Frågeställningar

Arbete I: Att utforska om variationen i deltagarnas kognitiva testresultat berodde på vilken testledare deltagarna hade träffat.

Arbete II: Att utforska kohortskillnader i deltagarnas kognitiva testresultat och om utbildningsnivån kunde förklara potentiella skillnader.

Urval från Populationen

För att besvara frågeställningarna i samtliga arbeten använde vi data från befolkningsstudien Gott Åldrande i Skåne (GÅS). Deltagare mellan 60 och 100 år var slumpmässigt utvalda från folkbokföringsregistret och kom till en av våra fyra mottagningar runt om Skåne. Deltagarna undersöktes av en läkare, en sjuksköterska och en psykologisk testledare och fick besvara frågor om livsstil och hälsa genom frågeformulär.

Metod och resultat

I arbete I (antal deltagare=6686) jämförde vi deltagarnas testresultat från 21 testledare i en flernivås-regressionsanalys. Vi kontrollerade för en rad faktorer som påverkar prestationen på kognitiva tester, såsom deltagarnas utbildning, ålder och testversion. Resultaten från analyserna visade en signifikant variation i deltagarnas testresultat (p<0.01), beroende av vilken testledare deltagaren hade träffat. Denna variation, som var mellan 1.4%-3.5%, förekom för kognitiva tester som mätte följande kognitiva funktioner: episodiskt minne, processhastighet och spatial förmåga.

I arbete II (antal 60 åringar =736, födda: 1942–43, 1948–49 eller 1954–55, antal 81 åringar= 431, födda 1920–21, 1926–27 eller 1932–33) jämförde vi deltagarnas testresultat, vilka var i samma ålder när de undersöktes (60 eller 81), men födda olika år, på kognitionstester som mätte en rad olika kognitiva funktioner. Resultaten från analyserna visade att de yngre kohorterna presterade bättre på kognitionstester som mätte processhastighet, episodiskt minne, semantiskt minne och exekutiv förmåga. Utbildningsnivån kunde inte förklara kohortskillnaderna i testresultaten.

Arbete III och IV lindrig kognitiv störning

Lindrig kognitiv störning (Eng. Mild Cognitive Impairment: MCI) är ett diagnostiskt begrepp som innebär att individen har en lätt kognitiv störning, men att störningen inte påverkar utförandet av vardagliga funktioner, såsom att handla mat eller klä på sig. MCI beskrivs som ett förstadium till demens. Prevalens, incidens och antal som har en tillbakagång av MCI (som inte längre anses möta MCI klassificeringen vid uppföljningen) varierar beroende på hur man definierar MCI. Just nu råder det ingen specifik konsensus kring vilka kriterier som ska ingå i MCI definitionen. De rådande prevalens- och incidenstalen i befolkningsstudier för individer över 60 år ligger på 18.9% och 47.9 per 1000 person-år (Petersen et al., 2014; Ward et al., 2012). En nyutkommen meta-analys (Petersen et al., 2018) rapporterade att drygt en fjärdedel (26.4%) av deltagarna, som anses ha MCI, i befolkningsstudier inte längre möter MCI kriterierna vid uppföljningsundersökningen. Anmärkningsvärt nog så varierar dessa siffror med skillnader utifrån studiernas design, populationsurvalet och MCI definitionen. Om MCI ska vara användbart som ett diagnostiskt verktyg, för att identifiera individer som kommer att utveckla demens, bör inte en stor andel visa på tillbakagång av MCI. För att kunna predicera vilka individer med MCI som har en god förutsättning att bli kognitivt bättre, bör forskningen fokusera på vilka faktorer som är associerade med tillbakagång. För tillfället är denna forskning knapphändig.

Frågeställning

Arbete III: Att rapportera prevalens och incidens av MCI, stratifierat för ålder, kön och subtyper, samt utforska om MCI estimaten varierade vid olika tillämpningar av MCI kriteriet för objektivt nedsatt kognition.

Arbete IV: Att rapportera tillbakagång av MCI efter 6år och undersöka de associerande faktorerna.

Metod och resultat

I både arbete III och IV användes den expanderade Mayo Clinic definitionen av MCI: 1) subjektivt eller objektivt rapporterad misstanke om nedsatt kognition. 2) ingen nedsatt förmåga i vardagliga aktiviteter. 3) minst ett nedsatt testresultat på följande kognitiva mått: episodiskt minne, processhastighet, spatial förmåga och verbalt flöde. 4) ingen demens (Petersen et al., 2014).

I arbete III (antal deltagare för prevalens=3752, antal deltagare för incidens=1451) rapporterades prevalens och incidens av MCI, uppdelat utifrån subtyper av MCI, kön och ålder. Vi använde två olika kriterier för MCI: ett lindrigt och ett strikt. För

att klassificeras som lindrigt MCI skulle deltagaren ha minst ett nedsatt testresultat på något av de kognitiva testerna och för det strikta kriteriet skulle deltagaren ha minst två i samma kognitiv domän. Vi följde friska deltagare (utan MCI) från deras första undersökning i ca 6år. Prevalensen för det lindriga kriteriet för MCI var 21.4% respektive 6.6% för det strikta kriteriet. De rapporterade incidenstalen var 22.6 och 8.87 per 1000 person-år för det lindriga kriteriet respektive det strikta kriteriet. Det fanns inga könsskillnader i prevalens eller incidens, men däremot en åldersskillnad i prevalens, där de yngsta och äldsta hade lägst respektive högst prevalens.

I arbete IV (antal deltagare = 331) undersöktes antalet med tillbakagång av MCI och vilka faktorer som var associerade med tillbakagång vid 6-årsuppföljningen. 58% av de deltagare som klassificerades som MCI vid första undersökningen ansågs friska, dvs. mötte inte kriterierna för MCI, vid 6-årsuppföljningen. Den logistiska regressionsmodellen upptäckte att lägre ålder ($p<0.05$), bättre global kognitiv förmåga ($p<0.02$), god koncentration ($p<0.05$) och nedsatt kognitiv förmåga i endast en kognitiv domän ($p<0.001$) (jämfört med flera kognitiva domäner) kunde predicera tillbakagång av MCI.

Sammanfattning och slutsats av samtliga arbeten

Sammantaget har den här avhandlingen fört fram bevis för testledareffekter och kohorteffekter på olika kognitiva testresultat. Den framhåller även att lindrig kognitiv störning, i ett urval av den äldre befolkningen, är relativt vanligt och att prevalens och incidens varierar med ålder och med MCI kriteriet man applicerar. Vidare har det presenterats att över hälften av deltagarna med MCI vid första undersökningen inte längre ansågs möta MCI kriterierna vid 6-årsundersökningen.

Denna avhandling visar på en betydelse av metodologiska aspekter vid utformandet och insamlandet av kognitiva data i den äldre befolkningen. Metodologiska beslut avseende på vilket sätt, var (hemma eller på mottagningen), vem (vilken testledare) och vilka ålderskohorter som ska undersökas kan påverka det kognitiva utfallet. Avhandlingen påvisar också att begreppet MCI måste förtydligas och definitionen preciseras, för att vara ett användbart diagnostiskt verktyg att predicera framtida demens med, i synnerhet vid applicerandet i befolkningsstudier.

Acknowledgements

The pleasure of reaching the finish line of this PhD journey is made complete by conveying my sincere gratitude.

First and foremost, I want to thank all the **GÅS-participants** for taking their time and providing us researchers with invaluable material.

To **Sölve Elmståhl**, my supervisor. Thank you for believing in me and giving me this opportunity. You have shared your knowledge, given me the freedom to form my research and encouraged me to face my weaknesses and therefore evolve as an independent researcher.

To my co-supervisor and friend **Mats Pihlsgård**. Your door has always been open to me. Besides providing invaluable and excellent statistical guidance you have patiently listened to my on- and off-topic thoughts and given considerate support and advice thereafter, for this I am utterly grateful. In the words of Serge pertaining my thesis- "*It's serious, it is a very important piece*".

To my friend and roommate **Vera Denvall**, you were the first to believe in me. Your expertise in neuropsychology has been instrumental, but also your advice, that being well organised can be as important as conducting the research itself. Thank you for all the deep conversations throughout the years.

To all the **test administrators** and the **psychologists** in training that I have worked with throughout the years, it has been a pleasure, and I have much appreciated our methodological discussions. But especially to you **Mia Lundström**, you have been a stable rock that has always cared about the core essence of data collection. I admire you and your enthusiasm, you have made this journey pleasurable. I owe you dinner!

To my existing and former colleagues in **Team I** and **II**, **Zinka, Irene, Johanna** and **Ole**. If it was not for your excellent efforts in organising, collecting, processing and storing data, us researchers would have nothing. Thank you for being lovely colleagues and keep up the good work!

To my colleagues and friends **Karin Werner** and **Helen Legrand**, you have made my time as a doctoral student exciting and enjoyable. I have always looked forward to poking my nose into your office for a laugh and a good conversation. You both inspire me to be a better researcher and person. Thank you, Helen, for your valuable

comments on paper III (and for your amazing cakes) and Karin for the delightful and thought-provoking lunches.

To the rest of my **colleagues** and fellow **PhD-students** at the Geriatric clinic, thank you for the stimulating conversations, for the comments on my thesis and for making an inspiring working environment. To **Henrik Ekström**, not only are you an admirable researcher but you also took your time to meticulously read through all my papers and thesis – a sincere thank you!

To my **family** and **friends**, your support and patience has helped me tremendously during this period. A special gratitude towards **Hanna** and **Louise** for making your time to read through my thesis, for your grammar control and insightful comments. A special appreciation to **Emilia** for sharing the hours of training that kept my brain thriving and for always being a fantastic listener.

I would also like to express appreciation towards the research school **SWEAH** for providing me with the tools to become a better researcher and teaching me the value of networking. To all my **SWEAH-colleagues**, this ride never felt lonely thanks to your good company.

To my love and husband **Joachim**. I can hardly express how grateful I am to have you in my life, especially during a time like this. Thank you for listening to my endless deliberations, for reading and providing insightful comments on this thesis, for taking Saga early in the mornings so I could rest or work and for being you. *– The best is yet to come*

References

Aerts, L., Heffernan, M., Kochan, N. A., Crawford, J. D., Draper, B., Trollor, J. N., . . . Brodaty, H. (2017). Effects of MCI subtype and reversion on progression to dementia in a community sample. *Neurology, 88*(23), 2225-2232. doi:10.1212/WNL.0000000000004015

Alwin, D. F., & McCammon, R. J. (2001). Aging, cohorts, and verbal ability. *J Gerontol B Psychol Sci Soc Sci, 56*(3), S151-161. doi:10.1093/geronb/56.3.s151

American Psychiatric Association, (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*: American Psychiatric Pub.

Annweiler, C., Fantino, B., Schott, A. M., Krolak-Salmon, P., Allali, G., & Beauchet, O. (2012). Vitamin D insufficiency and mild cognitive impairment: cross-sectional association. *Eur J Neurol, 19*(7), 1023-1029. doi:10.1111/j.1468-1331.2012.03675.x

Artero, S., Ancelin, M. L., Portet, F., Dupuy, A., Berr, C., Dartigues, J. F., . . . Ritchie, K. (2008). Risk profiles for mild cognitive impairment and progression to dementia are gender specific. *J Neurol Neurosurg Psychiatry, 79*(9), 979-984. doi:10.1136/jnnp.2007.136903

Au, B., Dale-McGrath, S., & Tierney, M. C. (2017). Sex differences in the prevalence and incidence of mild cognitive impairment: A meta-analysis. *Ageing Research Reviews, 35*, 176-199. doi:https://doi.org/10.1016/j.arr.2016.09.005

Belleville, S., Fouquet, C., Hudon, C., Zomahoun, H. T. V., & Croteau, J. (2017). Neuropsychological measures that predict progression from mild cognitive impairment to Alzheimer's type dementia in older adults: a systematic review and meta-analysis. *Neuropsychology Review, 27*(4), 328-353.

Belleville, S., Gauthier, S., Lepage, É., Kergoat, M.-J., & Gilbert, B. (2014). Predicting decline in mild cognitive impairment: A prospective cognitive study. *Neuropsychology, 28*(4), 643.

Bernstein, L. E., Tucker, P. E., & Auer, E. T., Jr. (1998). Potential perceptual bases for successful use of a vibrotactile speech perception aid. *Scand J Psychol, 39*(3), 181-186.

Blatter, K., & Cajochen, C. (2007). Circadian rhythms in cognitive performance: methodological constraints, protocols, theoretical underpinnings. *Physiol Behav, 90*(2-3), 196-208. doi:10.1016/j.physbeh.2006.09.009

Bondi, M. W., Edmonds, E. C., Jak, A. J., Clark, L. R., Delano-Wood, L., McDonald, C. R., . . . Salmon, D. P. (2014). Neuropsychological criteria for mild cognitive impairment improves diagnostic precision, biomarker associations, and progression rates. *J Alzheimers Dis, 42*(1), 275-289. doi:10.3233/jad-140276

Bookout, D. V. T., & Hosford, R. E. (1969). Administration effects on the S-329 of the Gatb using three experimental treatments. *Journal of Employment Counseling, 6*(3), 124-133. doi:10.1002/j.2161-1920.1969.tb00523.x

Brand, C. (1996). *The g factor: General intelligence and its implications*. Chichester, England: John Wiley & Sons.

Brodaty, H., Aerts, L., Crawford, J. D., Heffernan, M., Kochan, N. A., Reppermund, S., . . . Sachdev, P. S. (2017). Operationalizing the Diagnostic Criteria for Mild Cognitive Impairment: The Salience of Objective Measures in Predicting Incident Dementia. *Am J Geriatr Psychiatry, 25*(5), 485-497. doi:10.1016/j.jagp.2016.12.012

Brodaty, H., Heffernan, M., Kochan, N. A., Draper, B., Trollor, J. N., Reppermund, S., . . . Sachdev, P. S. (2013). Mild cognitive impairment in a community sample: The Sydney Memory and Ageing Study. *Alzheimer's & Dementia, 9*(3), 310-317.

Busse, A., Hensel, A., Gühne, U., Angermeyer, M., & Riedel-Heller, S. (2006). Mild cognitive impairment Long-term course of four clinical subtypes. *Neurology, 67*(12), 2176-2185.

Bäckman, L., Hill, R. D., & Forsell, Y. (1996). The influence of depressive symptomatology on episodic memory functioning among clinically nondepressed older adults. *Journal of abnormal psychology*, *105*(1), 97-105. doi:10.1037/0021-843X.105.1.97

Canevelli, M., Grande, G., Lacorte, E., Quarchioni, E., Cesari, M., Mariani, C., . . . Vanacore, N. (2016). Spontaneous Reversion of Mild Cognitive Impairment to Normal Cognition: A Systematic Review of Literature and Meta-Analysis. *Journal of the American Medical Directors Association, 17*(10), 943-948. doi:https://doi.org/10.1016/j.jamda.2016.06.020

Caracciolo, B., Palmer, K., Monastero, R., Winblad, B., Backman, L., & Fratiglioni, L. (2008). Occurrence of cognitive impairment and dementia in the community: a 9-year-long prospective study. *Neurology, 70*(19 Pt 2), 1778-1785. doi:10.1212/01.wnl.0000288180.21984.cb

Chapman, C. D., Benedict, C., & Schioth, H. B. (2018). Experimenter gender and replicability in science. *Sci Adv, 4*(1), e1701427. doi:10.1126/sciadv.1701427

Colman, A. M. (2006). *Oxford dictionary of psychology* (2nd ed.). New York, US: Oxford University Press.

Cosentino, S. A., Brickman, A. M., & Manly, J. J. (2011). Neuropsychological assessment of the dementias of late life *Handbook of the psychology of aging* (pp. 339-352): Elsevier.

Crook, T., Bartus, R. T., Ferris, S. H., Whitehouse, P., Cohen, G. D., & Gershon, S. (1986). Age-Associated Memory Impairment: Proposed Diagnostic Criteria and Measures of Clinical Change - Report of a National Institute of Mental Health Work Group. *Developmental Neuropsychology, 2*(4), 261-276. doi:10.1080/87565648609540348

Dahl, M., Allwood, C., Rennemark, M., & Hagberg, B. (2010). The relation between personality and the realism in confidence judgements in older adults. *European Journal of Ageing, 7*(4), 283-291. doi:10.1007/s10433-010-0164-2

Davies, K., Collerton, J. C., Jagger, C., Bond, J., Barker, S. A., Edwards, J., . . . Robinson, L. (2010). Engaging the oldest old in research: lessons from the Newcastle 85+ study. *BMC Geriatr, 10*, 64. doi:10.1186/1471-2318-10-64

Diaz-Mardomingo, M. D. C., Garcia-Herranz, S., Rodriguez-Fernandez, R., Venero, C., & Peraita, H. (2017). Problems in Classifying Mild Cognitive Impairment (MCI): One or Multiple Syndromes? *Brain Sci, 7*(9). doi:10.3390/brainsci7090111

Dickinson, M. D., & Hiscock, M. (2011). The Flynn effect in neuropsychological assessment. *Applied Neuropsychology, 18*(2), 136-142.

Dikmen, S. S., Heaton, R. K., Grant, I., & Temkin, N. R. (1999). Test-retest reliability and practice effects of expanded Halstead-Reitan Neuropsychological Test Battery. *J Int Neuropsychol Soc, 5*(4), 346-356.

Dodge, H. H., Zhu, J., Hughes, T. F., Snitz, B. E., Chang, C. H., Jacobsen, E. P., & Ganguli, M. (2017). Cohort effects in verbal memory function and practice effects: a population-based study. *Int Psychogeriatr, 29*(1), 137-148. doi:10.1017/S1041610216001551

Dodge, H. H., Zhu, J., Lee, C. W., Chang, C. C., & Ganguli, M. (2014). Cohort effects in age-associated cognitive trajectories. *J Gerontol A Biol Sci Med Sci, 69*(6), 687-694. doi:10.1093/gerona/glt181

Dutton, E., van der Linden, D., & Lynn, R. (2016). The negative Flynn Effect: A systematic literature review. *Intelligence, 59*, 163-169. doi:https://doi.org/10.1016/j.intell.2016.10.002

Dureman, I. (1960). *SRB:1*. Stockholm, Sweden: Psykologiförlaget.

Ekström, H., & Elmståhl, S. (2006). Pain and fractures are independently related to lower walking speed and grip strength: results from the population study "Good Ageing in Skåne". *Acta Orthopaedica, 77*(6), 902-911.

Elmståhl, S., & Widerström, E. (2014). Orthostatic intolerance predicts mild cognitive impairment: incidence of mild cognitive impairment and dementia from the Swedish general population cohort Good Aging in Skane. *Clin Interv Aging, 9*, 1993-2002. doi:10.2147/cia.s72316

Ferman, T., Smith, G., Pankratz, V. S., Kantarci, K., Boeve, B., Graff-Radford, N., . . . Petersen, R. (2014). Non-Amnestic Mild Cognitive Impairment Progresses to Dementia with Lewy Bodies. *Neurology*, *81*, 2032-2038. doi:10.1212/01.wnl.0000436942.55281.47

Ferrer, E., & Ghisletta, P. (2011). Methodological and analytical issues in the psychology of aging. *Handbook of the psychology of aging, 7*, 25-39.

Finkel, D., Reynolds, C. A., McArdle, J. J., & Pedersen, N. L. (2007). Cohort differences in trajectories of cognitive aging. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 62*(5), 286-294.

Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin, 95*(1), 29-51. doi:10.1037/0033-2909.95.1.29

Flynn, J. R., & Shayer, M. (2018). IQ decline and Piaget: Does the rot start at the top? *Intelligence, 66*, 112-121. doi:https://doi.org/10.1016/j.intell.2017.11.010

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). “Mini-mental state”: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*(3), 189-198. doi:http://dx.doi.org/10.1016/0022-3956(75)90026-6

Fratiglioni, L., Ding, M., Santoni, G., Berglund., Elmståhl, S., Fagerström, C., Lagergren, M., Sjölund. B. M., Sköldunger, A., Welmer, A. K., & Wimo, A. (2017). Demensförekomst i Sverige: geografiska trender 2001-2013. SNAC. Available online at: https://snacsweden.files.wordpress.com/2018/03/snac-rapport-demensfc3b6rekomst-i-sverige-geografiska-och-tidsmc3a4ssiga-trender-2001-2013.pdf

Galea, S., & Tracy, M. (2007). Participation rates in epidemiologic studies. *Ann Epidemiol, 17*(9), 643-653. doi:10.1016/j.annepidem.2007.03.013

Garand, L., Lingler, J. H., Conner, K. O., & Dew, M. A. (2009). Diagnostic labels, stigma, and participation in research related to dementia and mild cognitive impairment. *Res Gerontol Nurs, 2*(2), 112-121. doi:10.3928/19404921-20090401-04

Gardiner, J. M., & Java, R. I. (1993). Recognizing and remembering *Theories of memory* (pp. 163-188). Hove: Erlbaum.

Germine, L., Reinecke, K., & Chaytor, N. S. (2019). Digital neuropsychology: challenges and opportunities at the intersection of science and software. *The Clinical neuropsychologist, 33*(2), 271-286.

Gerstorf, D., Ram, N., Hoppmann, C., Willis, S. L., & Schaie, K. W. (2011). Cohort differences in cognitive aging and terminal decline in the Seattle Longitudinal Study. *Developmental Psychology, 47*(4), 1026-1041.

Giambra, L. M., Arenberg, D., Kawas, C., Zonderman, A. B., & Costa, P. T., Jr. (1995). Adult life span changes in immediate visual memory and verbal intelligence. *Psychol Aging, 10*(1), 123-139.

Gignac, G. E. (2015). The magical numbers 7 and 4 are resistant to the Flynn effect: No evidence for increases in forward or backward recall across 85 years of data. *Intelligence, 48*, 85-95.

Gillis, C., Mirzaei, F., Potashman, M., Ikram, M. A., & Maserejian, N. (2019). The incidence of mild cognitive impairment: A systematic review and data synthesis. *Alzheimers Dement (Amst), 11*, 248-256. doi:10.1016/j.dadm.2019.01.004

Goldberg, T. E., Harvey, P. D., Wesnes, K. A., Snyder, P. J., & Schneider, L. S. (2015). Practice effects due to serial cognitive assessment: Implications for preclinical Alzheimer's disease randomized controlled trials. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, 1*(1), 103-111. doi:http://dx.doi.org/10.1016/j.dadm.2014.11.003

Greenfield, P. M. (1998). The cultural evolution of IQ. In U. Neisser (Ed.), *The rising curve: Long-term gains in IQ and related measures* (pp. 81-123). Washington, DC, US: American Psychological Association. doi:/10.1037/10270-003.

Guaita, A., Vaccaro, R., Davin, A., Colombo, M., Vitali, S. F., Polito, L., . . . Villani, S. (2015). Influence of socio-demographic features and apolipoprotein E epsilon 4 expression on the prevalence of dementia and cognitive impairment in a population

of 70-74-year olds: the InveCe.Ab study. *Arch Gerontol Geriatr, 60*(2), 334-343. doi:10.1016/j.archger.2014.11.006

Guo, S., & Bollen, K. A. (2013). Research Using Longitudinal Ratings Collected by Multiple Raters: One Methodological Problem and Approaches to Its Solution. *Social Work Research, 37*(2), 85-98. doi:10.1093/swr/svs029

Gustafsson, J.-E. (2008). Schooling and intelligence: Effects of track of study on level and profile of cognitive abilities. In P.C. Kyllonen, R. D. Roberts, & L. Stankov (Eds.) *Extending intelligence: Enhancement and new constructs* (pp. 37-59). New York: Lawrence Erlbaum Associates.

Gwet, K. L. (2014). *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters*. Gaithersburgs, US: Advanced Analytics, LLC.

Hagberg, B., Bauer Alfredson, B., Poon, L. W., & Homma, A. (2001). Cognitive functioning in centenarians: a coordinated analysis of results from three countries. *J Gerontol B Psychol Sci Soc Sci, 56*(3), P141-151. doi:10.1093/geronb/56.3.p141

Han, J. W., Kim, T. H., Lee, S. B., Park, J. H., Lee, J. J., Huh, Y., . . . Kim, K. W. (2012). Predictive validity and diagnostic stability of mild cognitive impairment subtypes. *Alzheimer's & Dementia, 8*(6), 553-559. doi:10.1016/j.jalz.2011.08.007

Hanson, R., Smith, J. A., & Hume, W. (1985). Achievements of infants on items of the Griffiths scales: 1980 compared with 1950. *Child: care, health and development, 11*(2), 91-104.

Harada, C. N., Natelson Love, M. C., & Triebel, K. L. (2013). Normal cognitive aging. *Clin Geriatr Med, 29*(4), 737-752. doi:10.1016/j.cger.2013.07.002

Hertzog, C., Dixon, R. A., & Hultsch, D. F. (1992). Intraindividual change in text recall of the elderly. *Brain Lang, 42*(3), 248-269.

Hiscock, M. (2007). The Flynn effect and its relevance to neuropsychology. *Journal of Clinical & Experimental Neuropsychology, 29*(5), 514-529. doi:10.1080/13803390600813841

Hofer, S. M., & Alwin, D. F. (2008). *Handbook of cognitive aging: Interdisciplinary perspectives*. (Eds.) Thousand Oaks, US: Sage Publications.

Hogan, D. B., Jette, N., Fiest, K. M., Roberts, J. I., Pearson, D., Smith, E. E., . . . Maxwell, C. J. (2016). The Prevalence and Incidence of Frontotemporal Dementia: a Systematic Review. *Can J Neurol Sci, 43 Suppl 1*, S96-s109. doi:10.1017/cjn.2016.25

Hoyt, W. T. (2000). Rater bias in psychological research: When is it a problem and what can we do about it? *Psychological Methods, 5*(1), 64-86.

Hoyt, W. T., & Kerns, M. (1999). Magnitude and moderators of bias in observer ratings: A meta-analysis. *Psychological Methods, 4*(4), 403-424. doi:10.1037/1082-989X.4.4.403

Hughes, C. P., Berg, L., Danziger, W. L., Coben, L. A., & Martin, R. L. (1982). A new clinical scale for the staging of dementia. *Br J Psychiatry, 140*, 566-572.

Hugo, J., & Ganguli, M. (2014). Dementia and cognitive impairment: epidemiology, diagnosis, and treatment. *Clinics in geriatric medicine, 30*(3), 421-442. doi:10.1016/j.cger.2014.04.001

Hultsch, D. F. (1998). *Memory change in the aged*. New York, US: Cambridge University Press.

Jak, A. J., Preis, S. R., Beiser, A. S., Seshadri, S., Wolf, P. A., Bondi, M. W., & Au, R. (2016). Neuropsychological Criteria for Mild Cognitive Impairment and Dementia Risk in the Framingham Heart Study. *J Int Neuropsychol Soc, 22*(9), 937-943. doi:10.1017/s1355617716000199

Karlsson, P., Thorvaldsson, V., Skoog, I., Gudmundsson, P., & Johansson, B. (2015). Birth cohort differences in fluid cognition in old age: Comparisons of trends in levels and change trajectories over 30 years in three population-based samples. *Psychology and Aging, 30*(1), 83-94.

Katkov, M., Romani, S., & Tsodyks, M. (2014). Word length effect in free recall of randomly assembled word lists. *Front Comput Neurosci, 8*, 129. doi:10.3389/fncom.2014.00129

Katz, S., & Akpom, C. A. (1976). A measure of primary sociobiological functions. *Int J Health Serv, 6*(3), 493-508. doi:10.2190/UURL-2RYU-WRYD-EY3K

Koepsell, T. D., & Monsell, S. E. (2012). Reversion from mild cognitive impairment to normal or near-normal cognition: risk factors and prognosis. *Neurology, 79*(15), 1591-1598. doi:10.1212/WNL.0b013e31826e26b7

Kosteniuk, J. G., Morgan, D. G., O'Connell, M. E., Kirk, A., Crossley, M., Teare, G. F., . . . Quail, J. M. (2016). Simultaneous temporal trends in dementia incidence and prevalence, 2005-2013: a population-based retrospective cohort study in Saskatchewan, Canada. *Int Psychogeriatr, 28*(10), 1643-1658. doi:10.1017/S1041610216000818

Lagergren, M., Fratiglioni, L., Hallberg, I. R., Berglund, J., Elmstahl, S., Hagberg, B., . . . Wimo, A. (2004). A longitudinal study integrating population, care and social services data. The Swedish National study on Aging and Care (SNAC). *Aging Clin Exp Res, 16*(2), 158-168.

Larrieu, S., Letenneur, L., Orgogozo, J. M., Fabrigoule, C., Amieva, H., Le Carret, N., . . . Dartigues, J. F. (2002). Incidence and outcome of mild cognitive impairment in a population-based prospective cohort. *Neurology, 59*(10), 1594-1599. doi:10.1212/01.wnl.0000034176.07159.f8

Laukka, E. J., Lövdén, M., Herlitz, A., Karlsson, S., Ferencz, B., Pantzar, A., . . . Bäckman, L. (2013). Genetic effects on old-age cognitive functioning: A population-based study. *Psychology and Aging, 28*(1), 262-274. doi:10.1037/a0030829

Lee, S. B., Kim, K. W., Youn, J. C., Park, J. H., Lee, J. J., Kim, M. H., . . . Woo, J. I. (2009). Prevalence of Mild Cognitive Impairment and Its Subtypes Are Influenced by the Application of Diagnostic Criteria: Results from the Korean Longitudinal Study on Health and Aging (KLoSHA). *Dementia and Geriatric Cognitive Disorders, 28*(1), 23-29. doi:10.1159/000228841

Lezak, M. D., Howieson, D. B., & Loring, D. W. (2004). *Neuropsychological assessment* (4 ed.). New York, US: Oxford University Press.

Lim, G. S. (2011). The development and maintenance of rating quality in performance writing assessment: A longitudinal study of new and experienced raters. *Language Testing, 28*(4), 543–560. Doi:10.1177/0265532211406422

Lobo, A., LÓPez-AntÓN, R., De-La-CÁMara, C., Quintanilla, M. Á., Campayo, A., & Saz, P. (2008). Non-Cognitive Psychopathological Symptoms Associated with Incident Mild Cognitive Impairment and Dementia, Alzheimer's Type. *Neurotoxicity Research, 14*(2/3), 263-272.

Loewenstein, D. A., Acevedo, A., Small, B. J., Agron, J., Crocco, E., & Duara, R. (2009). Stability of different subtypes of mild cognitive impairment among the elderly over a 2- to 3-year follow-up period. *Dement Geriatr Cogn Disord, 27*(5), 418-423. doi:10.1159/000211803

Lopez, O. L., Becker, J. T., Chang, Y. F., Sweet, R. A., DeKosky, S. T., Gach, M. H., . . . Kuller, L. H. (2012). Incidence of mild cognitive impairment in the Pittsburgh Cardiovascular Health Study-Cognition Study. *Neurology, 79*(15), 1599-1606. doi:10.1212/WNL.0b013e31826e25f0

Lowe, C., & Rabbitt, P. (1998). Test/re-test reliability of the CANTAB and ISPOCD neuropsychological batteries: theoretical and practical issues. Cambridge Neuropsychological Test Automated Battery. International Study of Post-Operative Cognitive Dysfunction. *Neuropsychologia, 36*(9), 915-923.

Luck, T., Luppa, M., Briel, S., & Riedel-Heller, S. G. (2010). Incidence of mild cognitive impairment: a systematic review. *Dement Geriatr Cogn Disord, 29*(2), 164-175. doi:10.1159/000272424

Lynn, R. (2009). Fluid intelligence but not vocabulary has increased in Britain, 1979–2008. *Intelligence, 37*(3), 249-255.

Lövdén, M., Ghisletta, P., & Lindenberger, U. (2004). Cognition in the Berlin Aging Study (BASE): the first 10 years. *Aging Neuropsychology and Cognition, 11*(2-3), 104-133.

Malek-Ahmadi, M. (2016). Reversion From Mild Cognitive Impairment to Normal Cognition: A Meta-Analysis. *Alzheimer Dis Assoc Disord, 30*(4), 324-330. doi:10.1097/WAD.0000000000000145

Manly, J. J., Bell-McGinty, S., Tang, M. X., Schupf, N., Stern, Y., & Mayeux, R. (2005). Implementing diagnostic criteria and estimating frequency of mild cognitive impairment in an urban community. *Arch Neurol, 62*(11), 1739-1746. doi:10.1001/archneur.62.11.1739

Manly, J. J., Tang, M. X., Schupf, N., Stern, Y., Vonsattel, J. P., & Mayeux, R. (2008). Frequency and course of mild cognitive impairment in a multiethnic community. *Ann Neurol, 63*(4), 494-506. doi:10.1002/ana.21326

Marx, D. M., & Goff, P. A. (2005). Clearing the air: The effect of experimenter race on target's test performance and subjective experience. *British Journal of Social Psychology, 44*(4), 645-657. doi:10.1348/014466604X17948

May, C. P., Hasher, L., & Foong, N. (2005). Implicit memory, age, and time of day: paradoxical priming effects. *Psychol Sci, 16*(2), 96-100. doi:10.1111/j.0956-7976.2005.00788.x

McArdle, J. J., & Woodcock, R. W. (1997). Expanding test–retest designs to include developmental time-lag components. *Psychological Methods, 2*(4), 403-435. doi:10.1037/1082-989X.2.4.403

Mitchell, A. J., & Shiri-Feshki, M. (2009). Rate of progression of mild cognitive impairment to dementia – meta-analysis of 41 robust inception cohort studies. *Acta Psychiatrica Scandinavica, 119*(4), 252-265. doi:10.1111/j.1600-0447.2008.01326.x

Montgomery, S. A., & Asberg, M. (1979). A new depression scale designed to be sensitive to change. *The British journal of psychiatry, 134*(4), 382-389. doi:10.1192/bjp.134.4.382

Must, O., & Must, A. (2013). Changes in test-taking patterns over time. *Intelligence, 41*(6), 780-790. doi:10.1016/j.intell.2013.04.005

Nelson, A. P., & O'Connor, M. G. (2008). Mild cognitive impairment: a neuropsychological perspective. *CNS Spectrums: The International Journal of Neuropsychiatric Medicine, 13*(1), 56-64.

Nyberg, L., Lövdén, M., Riklund, K., Lindenberger, U., & Bäckman, L. (2012). Review: Memory aging and brain maintenance. *Trends in Cognitive Sciences, 16*, 292-305. doi:10.1016/j.tics.2012.04.005

Ortner, T. M., & Vormittag, I. (2011). Test administrator's gender affects female and male students' self-estimated verbal general knowledge. *Learning and Instruction, 21*(1), 14-21. doi:http://dx.doi.org/10.1016/j.learninstruc.2009.09.003

Palmer, K., Bäckman, L., Winblad, B., & Fratiglioni, L. (2008). Mild cognitive impairment in the general population: Occurrence and progression to Alzheimer disease. *The American Journal of Geriatric Psychiatry, 16*(7), 603-611. doi:10.1097/JGP.0b013e3181753a64

Pandya, S. Y., Clem, M. A., Silva, L. M., & Woon, F. L. (2016). Does mild cognitive impairment always lead to dementia? A review. *J Neurol Sci, 369*, 57-62. doi:10.1016/j.jns.2016.07.055

Pandya, S. Y., Lacritz, L. H., Weiner, M. F., Deschner, M., & Woon, F. L. (2017). Predictors of Reversion from Mild Cognitive Impairment to Normal Cognition. *Dement Geriatr Cogn Disord, 43*(3-4), 204-214. doi:10.1159/000456070

Pantzar, A., Laukka, E. J., Atti, A. R., Fastbom, J., Fratiglioni, L., & Bäckman, L. (2014). Cognitive deficits in unipolar old-age depression: a population-based study. *Psychological Medicine, 44*(5), 937-947. doi:10.1017/S0033291713001736

Park, M. H., & Han, C. (2015). Is there an MCI reversion to cognitively normal? Analysis of Alzheimer's disease biomarkers profiles. *International psychogeriatrics, 27*(3), 429-437.

Petersen, R. C. (2011). Mild cognitive impairment. *New England Journal of Medicine, 364*(23), 2227-2234.

Petersen, R. C., Caracciolo, B., Brayne, C., Gauthier, S., Jelic, V., & Fratiglioni, L. (2014). Mild cognitive impairment: a concept in evolution. *Journal of internal medicine, 275*(3), 214-228. doi:10.1111/joim.12190

Petersen, R. C., Lopez, O., Armstrong, M. J., Getchius, T. S. D., Ganguli, M., Gloss, D., . . . Rae-Grant, A. (2018). Practice guideline update summary: Mild cognitive impairment: Report of the Guideline Development, Dissemination, and Implementation Subcommittee of the American Academy of Neurology. *Neurology, 90*(3), 126-135. doi:10.1212/WNL.0000000000004826

Petersen, R. C., Roberts, R. O., Knopman, D. S., Boeve, B. F., Geda, Y. E., Ivnik, R. J., . . . Jack, C. R. (2009). Mild cognitive impairment: ten years later. *Archives of neurology, 66*(12), 1447-1455.

Petersen, R. C., Roberts, R. O., Knopman, D. S., Geda, Y. E., Cha, R. H., Pankratz, V. S., . . . Rocca, W. A. (2010). Prevalence of mild cognitive impairment is higher in men. The Mayo Clinic Study of Aging. *Neurology, 75*(10), 889-897. doi:10.1212/WNL.0b013e3181f11d85

Petersen, R. C., Smith, G. E., Waring, S. C., Ivnik, R. J., Tangalos, E. G., & Kokmen, E. (1999). Mild cognitive impairment: Clinical characterization and outcome. *Archives of neurology, 56*(3), 303-308. doi:10.1001/archneur.56.3.303

Pietschnig, J., & Voracek, M. (2015). One Century of Global IQ Gains: A Formal Meta-Analysis of the Flynn Effect (1909-2013). *Perspect Psychol Sci, 10*(3), 282-306. doi:10.1177/1745691615577701

Prince, M., Bryce, R., Albanese, E., Wimo, A., Ribeiro, W., & Ferri, C. P. (2013). The global prevalence of dementia: a systematic review and metaanalysis. *Alzheimer's & Dementia, 9*(1), 63-75. e62.

Prince, M., Lewis, G., Bird, A., Blizard, R., & Mann, A. (1996). A longitudinal study of factors predicting change in cognitive test scores over time, in an older hypertensive population. *Psychol Med, 26*(3), 555-568.

Ravaglia, G., Forti, P., Montesi, F., Lucicesare, A., Pisacane, N., Rietti, E., . . . Mecocci, P. (2008). Mild cognitive impairment: epidemiology and dementia risk in an elderly Italian population. *J Am Geriatr Soc, 56*(1), 51-58. doi:10.1111/j.1532-5415.2007.01503.x

Rehnman, J., & Herlitz, A. (2006). Higher face recognition ability in girls: Magnified by own-sex and own-ethnicity bias. *Memory, 14*(3), 289-296. doi:10.1080/09658210500233581

Reitan, R. M., & Wolfson, D. (1985). *The Halstead-Reitan neuropsychological test battery: Theory and clinical interpretation* (Vol. 4): Reitan Neuropsychology.

Rizzi, L., Rosset, I., & Roriz-Cruz, M. (2014). Global epidemiology of dementia: Alzheimer's and vascular types. *BioMed research international, 2014*.

Roberts, R. O., Knopman, D. S., Mielke, M. M., Cha, R. H., Pankratz, V. S., Christianson, T. J., . . . Petersen, R. C. (2014). Higher risk of progression to dementia in mild cognitive impairment cases who revert to normal. *Neurology, 82*(4), 317-325. doi:10.1212/WNL.0000000000000055

Roebuck-Spencer, T. M., Glen, T., Puente, A. E., Denney, R. L., Ruff, R. M., Hostetter, G., & Bianchini, K. J. (2017). Cognitive Screening Tests Versus Comprehensive Neuropsychological Test Batteries: A National Academy of Neuropsychology Education Paper†. *Archives of Clinical Neuropsychology, 32*(4), 491-498. doi:10.1093/arclin/acx021

Rosenthal, R., & Fode, K. L. (1963). Psychology of the scientist: V. Three experiments in experimenter bias. Monograph Supplement 3-V12. *Psychological Reports, 12*(2), 491-511.

Rothman, K., & Hatch, E. (2002). *An introduction: epidemiology*. New York, US: Oxford University Press

Rousson, V., Gasser, T., & Seifert, B. (2002). Assessing intrarater, interrater and test-retest reliability of continuous measurements. *Stat Med, 21*(22), 3431-3446. doi:10.1002/sim.1253

Rönnlund, M., & Nilsson, L.-G. (2008). The magnitude, generality, and determinants of Flynn effects on forms of declarative memory and visuospatial ability: Time-sequential analyses of data from a Swedish cohort study. *Intelligence, 36*(3), 192-209. doi:http://dx.doi.org/10.1016/j.intell.2007.05.002

Sachdev, P. S., Lipnicki, D. M., Crawford, J., Reppermund, S., Kochan, N. A., Trollor, J. N., . . . Sydney Memory, A. S. T. (2013). Factors predicting reversion from mild cognitive impairment to normal cognitive functioning: a population-based study. *PLoS One, 8*(3), e59649. doi:10.1371/journal.pone.0059649

Salthouse, T. A. (2010). Influence of age on practice effects in longitudinal neurocognitive change. *Neuropsychology, 24*(5), 563-572. doi:10.1037/a0019026

Salthouse, T. A. (2014). Aging Cognition Unconfounded by Prior Test Experience. *J Gerontol B Psychol Sci Soc Sci*. doi:10.1093/geronb/gbu063

Salthouse, T. A. (2015). Implications of the Flynn effect for age–cognition relations. *Intelligence, 48*, 51-57.

Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology, 27*(5), 763-776. doi:10.1037/0012-1649.27.5.763

Samuel, W. (1977). Observed IQ as a function of test atmosphere, tester expectation, and race of tester: A replication for female subjects. *Journal of Educational Psychology, 69*(5), 593-604. doi:10.1037/0022-0663.69.5.593

Sattler, J. M., & Theye, F. (1967). Procedural, situational, and interpersonal variables in individual intelligence testing. *Psychological Bulletin, 68*(5), 347-360. doi:10.1037/h0025153

Schaie, K. W. (1988). Internal Validity Threats in Studies of Adult Cognitive Development *Cognitive Development in Adulthood*. New York: Springer.

Schaie, K. W. (2011). Historical influences on aging and behavior. *Handbook of the psychology of aging, 7*, 41-55.

Schaie, K. W., & Willis, S. L. (2010). *Handbook of the Psychology of Aging*:Academic Press.

Schmidt, C., Collette, F., Cajochen, C., & Peigneux, P. (2007). A time to think: circadian rhythms in human cognition. *Cognitive Neuropsychology, 24*(7), 755-789.

Schmidt, K. R. T., & Teti, D. M. (2005). Issues in the use of longitudinal and cross-sectional designs. *Handbook of research methods in developmental science*, 3-20.

Seblova, D., Quiroga, M. L., Fors, S., Johnell, K., Lovden, M., de Leon, A. P., . . . Lager, A. (2018). Thirty-year trends in dementia: a nationwide population study of Swedish inpatient records. *Clin Epidemiol, 10*, 1679-1693. doi:10.2147/CLEP.S178955

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA, US: Houghton, Mifflin and Company

Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer* (Vol. 1), Newbury Park: Sage.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171*(3972), 701-703.

Sherwood, B., Zhou, A. X., Weintraub, S., & Wang, L. (2016). Using quantile regression to create baseline norms for neuropsychological tests. *Alzheimers Dement (Amst), 2*, 12-18. doi:10.1016/j.dadm.2015.11.005

Shievitz, A. L., Tudiver, F., Araujo, A., Sanghe, P., & Boyle, E. (1998). Do elderly people score better on cognitive tests at home? *Canadian Family Physician Médecin De Famille Canadien, 44*, 1652-1656.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychol Bull, 86*(2), 420-428.

Skirbekk, V., Stonawski, M., Bonsang, E., & Staudinger, U. M. (2013). The Flynn effect and population aging. *Intelligence, 41*(3), 169-177.

Sliwinski, M. J., Mogle, J. A., Hyun, J., Munoz, E., Smyth, J. M., & Lipton, R. B. (2018). Reliability and Validity of Ambulatory Cognitive Assessments. *Assessment, 25*(1), 14-30. doi:10.1177/1073191116643164

Snaith, R., Harrop, F., Newby, D., & Teale, C. (1986). Grade scores of the Montgomery-Åsberg Depression and the Clinical Anxiety Scales. *The British journal of psychiatry*(148), 599-601.

Solfrizzi, V., Panza, F., Colacicco, A. M., D'Introno, A., Capurso, C., Torres, F., . . . Italian Longitudinal Study on Aging Working, G. (2004). Vascular risk factors, incidence of MCI, and rates of progression to dementia. *Neurology, 63*(10), 1882-1891. doi:10.1212/01.wnl.0000144281.38555.e3

Strauss, H., Sherman, E., & Spreen, O. (2006). *A Compendium of Neuropsychological Tests: Administration, Norms and Commentary* (Vol. 2nd ed). New York, US: Oxford University Press.

Sugarman, M. A., Alosco, M. L., Tripodis, Y., Steinberg, E. G., & Stern, R. A. (2018). Neuropsychiatric symptoms and the diagnostic stability of mild cognitive impairment. *Journal of Alzheimer's Disease, 62*(4), 1841-1855.

Summers, M. J., & Saunders, N. L. J. (2012). Neuropsychological measures predict decline to Alzheimer's dementia from mild cognitive impairment. *Neuropsychology, 26*(4), 498-508. doi:10.1037/a0028576

Sundet, J. M., Barlaug, D. G., & Torjussen, T. M. (2004). The end of the Flynn effect?: A study of secular trends in mean intelligence test scores of Norwegian conscripts during half a century. *Intelligence, 32*(4), 349-362.

Teasdale, T. W., & Owen, D. R. (2008). Secular declines in cognitive test scores: A reversal of the Flynn Effect. *Intelligence, 36*(2), 121-126.

Tervo, S., Kivipelto, M., Hänninen, T., Vanhanen, M., Hallikainen, M., Mannermaa, A. (2004). Incidence and risk factors for mild cognitive impairment: a population-based three-year follow-up study of cognitively healthy elderly subjects. *Dement Geriatr Cogn Disord*, 17(3), 196–203.

Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology, 4*(1), 25-29. doi:10.1037/h0071663

Thorvaldsson, V., Karlsson, P., Skoog, J., Skoog, I., & Johansson, B. (2017). Better Cognition in New Birth Cohorts of 70 Year Olds, But Greater Decline Thereafter. *J Gerontol B Psychol Sci Soc Sci, 72*(1), 16-24. doi:10.1093/geronb/gbw125

Tokuchi, R., Hishikawa, N., Kurata, T., Sato, K., Kono, S., Yamashita, T., . . . Abe, K. (2014). Clinical and demographic predictors of mild cognitive impairment for converting to Alzheimer's disease and reverting to normal cognition. *J Neurol Sci, 346*(1-2), 288-292. doi:10.1016/j.jns.2014.09.012

Tombaugh, T. N., Kozak, J., & Rees, L. (1999). Normative data stratified by age and education for two measures of verbal fluency: FAS and animal naming. *Arch Clin Neuropsychol, 14*(2), 167-177.

Tombaugh, T. N., & McIntyre, N. J. (1992). The mini-mental state examination: a comprehensive review. *J Am Geriatr Soc, 40*(9), 922-935. doi:10.1111/j.1532-5415.1992.tb01992.x

Trahan, L. H., Stuebing, K. K., Fletcher, J. M., & Hiscock, M. (2014). The Flynn effect: A meta-analysis. *Psychological Bulletin, 140*(5), 1332-1360. doi:10.1037/a0037173

Trittschuh, E. H., Crane, P. K., Larson, E. B., Cholerton, B., McCormick, W. C., McCurry, S. M., . . . Craft, S. (2011). Effects of varying diagnostic criteria on prevalence of mild cognitive impairment in a community based sample. *J Alzheimers Dis, 25*(1), 163-173. doi:10.3233/jad-2011-101821

Tuddenham, R. D. (1948). Soldier intelligence in World Wars I and II. *Am Psychol, 3*(2), 54-56.

Twisk, J. W. (2006). *Applied multilevel analysis: a practical guide for medical researchers*: Cambridge, UK: Cambridge university press.

United Nations Department of Economic and Social Affairs Population Division (2017). World Population Ageing 2017 – Highlights (ST/ESA/SER.A/397). Available online at: http://www.un.org/en/development/desa/population/publications/pdf/ageing/WPA2017_Highlights.pdf

Van Beijsterveldt, C. E., van Boxtel, M. P., Bosma, H., Houx, P. J., Buntinx, F., & Jolles, J. (2002). Predictors of attrition in a longitudinal cognitive aging study: the Maastricht Aging Study (MAAS). *J Clin Epidemiol, 55*(3), 216-223.

Verhaeghen, P. (2013). *The elements of cognitive aging: Meta-analyses of age-related differences in processing speed and their consequences*. New York, US: Oxford University Press.

Visser, P. J., Verhey, F., Knol, D. L., Scheltens, P., Wahlund, L.-O., Freund-Levi, Y., . . . Hampel, H. (2009). Prevalence and prognostic value of CSF markers of Alzheimer's disease pathology in patients with subjective cognitive impairment or mild cognitive impairment in the DESCRIPA study: a prospective cohort study. *The Lancet Neurology, 8*(7), 619-627.

Ward, A., Arrighi, H. M., Michels, S., & Cedarbaum, J. M. (2012). Mild cognitive impairment: disparity of incidence and prevalence estimates. *Alzheimers Dement, 8*(1), 14-21. doi:10.1016/j.jalz.2011.01.002

Watson, F. L., Pasteur, M.-A. L., Healy, D. T., & Hughes, E. A. (1994). Nine Parallel Versions of Four Memory Tests: An Assessment of Form Equivalence and the Effects of Practice on Performance. *Human Psychopharmacology: Clinical & Experimental, 9*(1), 51-61. doi:10.1002/hup.470090107

Wechsler, D. (1981). *WAIS-R manual: Wechsler adult intelligence scale-revised*: Psychological Corporation.

Williams, R. L. (2013). Overview of the Flynn effect. *Intelligence, 41*(6), 753-764. doi:http://dx.doi.org/10.1016/j.intell.2013.04.010

Wilson, R. S., Beckett, L. A., Barnes, L. L., Schneider, J. A., Bach, J., Evans, D. A., & Bennett, D. A. (2002). Individual differences in rates of change in cognitive abilities of older persons. *Psychology and Aging, 17*(2), 179-193. doi:10.1037/0882-7974.17.2.179

Wimo, A., Guerchet, M., Ali, G.-C., Wu, Y.-T., Prina, A. M., Winblad, B., . . . Prince, M. (2017). The worldwide costs of dementia 2015 and comparisons with 2010. *Alzheimer's & Dementia, 13*(1), 1-7.

Winblad, B., Amouyel, P., Andrieu, S., Ballard, C., Brayne, C., Brodaty, H., . . . Feldman, H. (2016). Defeating Alzheimer's disease and other dementias: a priority for European science and society. *The Lancet Neurology, 15*(5), 455-532.

Winblad, B., Palmer, K., Kivipelto, M., Jelic, V., Fratiglioni, L., Wahlund, L. O., . . . Almkvist, O. (2004). Mild cognitive impairment–beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment. *Journal of internal medicine, 256*(3), 240-246.

Woodford, H. J., & George, J. (2007). Cognitive assessment in the elderly: a review of clinical methods. *QJM, 100*(8), 469-484. doi:10.1093/qjmed/hcm051

World Health Organization.World Health Statistics 2018: Monitor-ing Health for the SDGs, Sustainable Development Goals (2018). Geneva:World Health Organization. Avaliable online at: https://www.who.int/gho/publications/world_health_statistics/2016/en/

Zazzo, R. (1974). Test des deux barrages. Actualités pédagogiques et psychologiques. Vol. 7. Neuchâtel, Switzerland: Delachaux et Nestlé.

Zelinski, E. M., & Kennison, R. F. (2007). Not your parents' test scores: cohort reduces psychometric aging effects. *Psychology and Aging, 22*(3), 546-557. doi:10.1037/0882-7974.22.3.546


Somewhere
something
incredible is
waiting to
be known.