<table>
<thead>
<tr>
<th>Chapter</th>
<th>Type of study</th>
<th>Disease field</th>
<th>Type of data</th>
<th>Predictors or determinant in final model</th>
<th>Outcome</th>
<th>Main findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Prediction:  development, internal validation, external validation</td>
<td>Type 2 diabetes</td>
<td>4 population-based cohorts</td>
<td>Age, BMI, waist circumference, use of anti-hypertensives, smoking, parental history of diabetes</td>
<td>HbA1c levels</td>
<td>Moderate model performance</td>
</tr>
<tr>
<td>3</td>
<td>Prediction: external validation, updating</td>
<td>Type 2 diabetes</td>
<td>4 population-based cohorts</td>
<td>Age, BMI, waist circumference, use of antihypertensives, smoking, family history of diabetes</td>
<td>Developing CVD, T2D or CKD</td>
<td>Adequate model performance after update of model intercept</td>
</tr>
<tr>
<td>4</td>
<td>Prediction: external validation, updating</td>
<td>Pneumonia in nursing home residents</td>
<td>5 nursing home cohorts</td>
<td>Gender, respiratory rate, respiratory difficulty, pulse rate, decreased alertness, insufficient fluid intake, eating dependency, pressure sore</td>
<td>Mortality</td>
<td>Adequate model performance after update of model intercept</td>
</tr>
<tr>
<td>5</td>
<td>Prediction: external validation, updating, internal validation</td>
<td>Pneumonia in nursing home residents</td>
<td>nursing home cohort</td>
<td>Gender, respiratory rate, respiratory difficulty, pulse rate, decreased alertness, insufficient fluid intake, eating dependency, dehydration, bowel incontinence, increase in eating dependency, cardiovascular history</td>
<td>Mortality</td>
<td>Adequate model performance; performance improved after model revision and model extension</td>
</tr>
<tr>
<td>6</td>
<td>Etiological (cross-sectional data from prospective cohort)</td>
<td>Type 2 diabetes</td>
<td>population-based cohort</td>
<td>Social jetlag</td>
<td>Cardiometabolic risk factors / T2D</td>
<td>Social jetlag was associated with cardiometabolic risk factors; For younger people: also association with T2D</td>
</tr>
<tr>
<td>7</td>
<td>Etiological (prospective)</td>
<td>Type 2 diabetes</td>
<td>cohort using routinely collected data (T2D population)</td>
<td>Hypoglycaemia</td>
<td>Mortality</td>
<td>Mild hypoglycaemia was associated with a lower mortality rate; severe hypoglycaemia requiring medical assistance was non-significantly associated with higher mortality</td>
</tr>
<tr>
<td>8</td>
<td>Etiological (cross-sectional data from prospective cohort)</td>
<td>Type 2 diabetes</td>
<td>cohort using routinely collected data + questionnaires (T2D population)</td>
<td>Level of occupation of individual and partner</td>
<td>HbA1c levels</td>
<td>For women, an unemployed partner was associated with higher HbA1c levels; for men, a partner with intermediate level of occupation was associated with lower HbA1c levels</td>
</tr>
</tbody>
</table>

Abbreviations: BMI: body mass index; CKD: chronic kidney disease; CVD: cardiovascular disease; T2D: type 2 diabetes
The main objective of this thesis was to develop, validate and update prediction models and to identify potential etiological factors in epidemiological cohort studies in the field of type 2 diabetes and in the field of pneumonia in nursing home residents with dementia. The aim of this thesis was to apply existing methods for prediction modelling and for etiological modelling to these fields.

In this chapter, first, the main findings of the previous chapters are summarized and interpreted in relation to the literature. Table 1 provides an overview of the different chapters of this thesis, including the type of study, the disease field, the type of data that was used for the study, the predictors that were included in the final prediction model, or, –for the etiological studies– the determinant under study, and the main findings.

Second, this chapter focuses on methodological issues that were encountered in the application of these methods, with regard to combining cohort studies, confounding in etiological models versus prediction models, using routinely collected data for research purposes, and applying the same model in different populations. Finally, in this chapter, implications and recommendations are provided for clinical practice, research practice and future research.

**MAIN FINDINGS AND REFLECTIONS**

**Predicting type 2 diabetes**

In chapter 2, we developed the DIRECT-DETECT prediction model, a model for predicting HbA1c levels after six years of follow-up in a non-diabetic general population. In order to facilitate the use of the model in clinical practice and –for research purposes– in large databases where blood assays are not available, we only included predictors in the model that can be measured non-invasively, and thus, laboratory-based predictors were not included (Table 1). Earlier studies used non-invasive predictors to predict the risk of developing type 2 diabetes [1-12], but none of these included HbA1c levels as a diagnostic criterion to define type 2 diabetes. We were the first to show that these predictors can also be used to predict HbA1c levels. However, the performance of the model was moderate: the model systematically overestimated HbA1c levels at follow-up in two out of the three cohorts that were used to develop the model. In addition, the model systematically underestimated HbA1c levels in the third development cohort as well as in the cohort that was used to externally validate the model. Possibly, this was due to the fact that different assays were used between the cohort studies to measure HbA1c levels. This indicates that the model's intercept should be adjusted for each cohort to improve predictions. Moreover, these predictors could only explain a small part of the observed variance in HbA1c levels, and the discriminative performance of the model was moderate, limiting its use as a screening tool in clinical practice.
In chapter 3, we externally validated another non-invasive risk prediction model that was previously developed for predicting the combined 7-year risk for chronic cardiometabolic diseases, including type 2 diabetes, cardiovascular disease, and chronic kidney disease [13]. We observed that the discriminative ability of the model was acceptable in The Australian Diabetes, Obesity and Lifestyle Study. This indicates that the predictors that were found to be relevant predictors in the development study, were also relevant for this different population, and also the same predictor estimates could be applied. Thus, the model can adequately discriminate between people who have relatively low or relatively high risk to develop these diseases. In addition, calibration of the model was adequate after updating the model intercept. The results of this external validation study add strength to the validity of the risk prediction model in different populations. This way, the risk prediction model is a useful tool in large-scale prevention programmes as a first step in the identification of individuals who are in need of further risk factor profiling and who might benefit from interventions.

Predicting mortality in nursing home residents with dementia and pneumonia

In chapter 4, we evaluated whether a previously developed model for predicting 14-day mortality in nursing home residents with dementia and pneumonia treated with antibiotics [14] could be applied to residents who were not treated with antibiotics. This model performed well in the development dataset in terms of discrimination and calibration. This indicates that the predictors that were found to be relevant predictors of short-term mortality for residents treated with antibiotics, were also relevant for the untreated residents, and also the same predictor estimates can be applied. Thus, the model can adequately discriminate between untreated residents who have relatively low or high risk to die within 14 days. In addition, calibration of the model was adequate after updating the model intercept. The updated model can therefore be a useful tool to predict short-term mortality risk in nursing home residents with dementia and pneumonia not treated with antibiotics.

In chapter 5, we evaluated the performance of the same prediction model for treated residents –developed in a cohort from the 1990s– in a more recent cohort (~20 years later). Since the development of the model, care for older people changed, as have population characteristics of nursing home residents. These changes include improved nutritional and hydration status, as well as improved overall health condition and lower 14-day mortality. We observed that the performance of the model was still adequate, although the discriminative performance of the model was slightly lower in our more recent cohort. Therefore, we updated the model, by evaluating whether predictors should be revised and whether additional predictors should be added to the model to improve the discriminative performance of the model. We found that the predictor ‘pressure sore’, which was only borderline significant in the development cohort, was no longer significant in the more recent cohort and could now be removed from the model, while
the following predictors were added –related to the above-mentioned changes in care and in resident characteristics: dehydration, bowel incontinence, cardiovascular history and increase in eating dependency. This updating procedure indeed improved model performance, leading to similar discriminative performance as the original model in the 1990s cohort.

**Etiological models in the field of type 2 diabetes**

In chapter 6, we evaluated the association between social jetlag, cardiometabolic risk factors and type 2 diabetes in cross-sectional data from the New Hoorn Study cohort. Social jetlag was defined as the difference in midpoint sleep –that is, the midpoint between bedtime and wake time– between weekdays and weekend days. For this association, we observed effect modification by age: in older people, no significant associations were found between social jetlag, cardiometabolic risk factors and diabetes/prediabetes. This might be explained by the lower prevalence of social jetlag in this group, possibly due to the absence of work obligations in higher age. As a result, the ‘no social jetlag’ group might be a mix of people who never had social jetlag and people who had social jetlag for more than 40 years. However, in younger people, a social jetlag of >2h was significantly associated with an increased prevalence of cardiometabolic risk factors and diabetes/prediabetes, compared to no social jetlag.

We adjusted the analyses for several demographic, clinical and behavioural characteristics that could confound this association. As this was cross-sectional research, no difference could be made between possible confounding factors, or possible intermediate factors (i.e., factors that are part of the causal chain between the determinant of interest and the outcome). Therefore, instead of combining all potential confounders in one adjusted model, separate models were presented, allowing for interpretation of the effect of adjusting for the separate factors. For example, adjusting for body mass index reduced the size of the associations. In addition, waist circumference was one of the cardiometabolic risk factors that was significantly increased in participants with a larger social jetlag. Although research in prospective cohorts is necessary to confirm any hypothesis, based on these results, we hypothesize that visceral obesity is a mediator between social jetlag, cardiometabolic risk factors and type 2 diabetes.

In chapter 7, we evaluated the association between self-reported hypoglycaemia and mortality. We observed that people who reported only mild hypoglycaemia had a significantly lower mortality rate during a median follow-up of 2 years, compared with people who reported no hypoglycaemia, and adjusting for demographic and clinical characteristics. This finding was in line with two previous studies on objectively measured [15] and self-reported mild hypoglycaemia [16], but contradictory to two other studies, which showed (non-significant) higher mortality rates for objectively measured [17] and self-reported mild hypoglycaemia [18]. Discrepancies between the results of these studies...
might be explained by differences between the study populations regarding diabetes duration, type of diabetes treatment, type of diabetes, and care setting.

Surprisingly, in contrast with previous studies [15-23], we also observed a lower mortality rate in people who reported severe hypoglycaemia, compared with people who reported no hypoglycaemia, although this difference was not significant. To further explore this result, we performed several sensitivity analyses. These showed that reporting severe hypoglycaemia not requiring medical help was non-significantly associated with a lower mortality rate, while reporting severe hypoglycaemia requiring medical assistance was non-significantly associated with a higher mortality rate. The results of these sensitivity analyses helped to explain the results that seemed surprising at first sight. Still, we were hesitant to conclude a causal relation between hypoglycaemia and a lower mortality rate. For example, our results could be explained by the concept of impaired awareness of hypoglycaemia [24]: people who reported no hypoglycaemia might in fact have an impaired awareness of hypoglycaemia, which could prevent them from taking actions to resolve their hypoglycaemic events. Unfortunately, we did not have measures of hypoglycaemia awareness and therefore more research is needed to verify this possible mechanism. As a results of these limitations, we formulated our conclusion rather conservative, concluding that people reporting hypoglycaemia not requiring medical assistance did not have an increased risk of mortality, suggesting that these sensations are not an indicator of increased short-term mortality risk in people with type 2 diabetes.

In chapter 8, we evaluated the association between individual and partner’s level of occupation on HbA1c levels in people with type 2 diabetes in cross-sectional data from the Diabetes Pearl cohort. We found that these associations differed between women and men: for women, having an unemployed partner was associated with higher HbA1c levels, compared with having a partner of the highest occupational level, and adjusted for demographic characteristics, diabetes medication, and recruitment center. This association was in line with previous studies where lower partner socio-economic status was associated with worse health outcomes [25, 26]. In contrast, for the men in our study, having a partner with an intermediate level of occupation was significantly associated with lower HbA1c levels, compared with having a partner of the highest occupational level. Underlying pathways of the association between partner socio-economic status and health outcomes are not yet clear and could not be evaluated in our study. Our study however does indicate that partner’s occupational level provides important information, and more research is necessary to explore possible underlying pathways.

**METHODOLOGICAL ISSUES**

During prediction modelling and etiological modelling in the field of type 2 diabetes and the field of pneumonia in nursing home residents with dementia, we encountered several
methodological issues. The next part specifically focuses on the following methodological issues. First, how to develop etiological and prediction models using combined individual patient data from different cohort studies? Second, what is the role of confounding in prediction models compared to etiological models? Third, how to use cohorts consisting of data that were routinely collected in clinical practice? And fourth, how to use the same prediction model in different populations, even when the population is treated differently? When is updating required?

Combining cohort studies

Meta-analyses of published studies and meta-analyses of individual patient data are widely used methods to combine results from previously conducted studies [27]. In several chapters of this thesis, we combined (i.e., pooled) individual patient data from different cohort studies. Compared to analyzing data from a single study, pooled analyses can be expected to increase the precision and generalizability of the results as the results are derived using a broader population [28, 29]. Indeed, when we externally validated the model for predicting the combined 7-year risk for chronic cardiometabolic diseases (Ch. 3), which was previously developed by pooled analyses of three cohort studies, we observed only a small decrease in discriminative performance in our external validation dataset compared to that in the development dataset. This indicates that overestimation of model performance, which is often reported when assessing performance in the development data [30], was scarcely present in the development study. This may partly be explained by the pooled analyses that were performed during the development of the model, which may have contributed to robust performance across populations [28, 29].

Another advantage of pooling individual patient data is that this increases the sample size. First, when we externally validated the prediction model for untreated pneumonia (Ch. 4), the number of untreated nursing home residents was very low in the available cohorts, and external validation of the model was possible only through pooled analyses. Second, when we pooled data from different cohorts to develop the DIRECT-DETECT prediction model (Ch. 2), the increased sample size may have reduced the risk of overoptimistic estimates, as overoptimism is particularly a problem in small datasets [30]. Third, when we combined data from people with type 2 diabetes treated in primary, secondary and tertiary care into one cohort study, namely the Diabetes Pearl cohort, which was used to evaluate the association between level of occupation and HbA1c levels (Ch. 8), the increased sample size improved the power to detect relevant relations.

One methodological challenge that we encountered when pooling individual patient data is related to differences in the definition and availability of variables between existing cohorts. Whereas in the Diabetes Pearl cohort (Ch. 8), the same procedures were used to prospectively collect data in the participating centers to ensure comparability of the data [31], in other chapters comparability of the data was sometimes a challenge when combining data from existing cohorts. For example, when we developed the DIRECT-
DETECT prediction model (Ch. 2), one of the predictors of interest was ‘family history of diabetes’. However, in one of the three datasets that was used to develop the model, no information was available on the diabetes status of siblings. Therefore, we used ‘parental history of diabetes’ as a proxy predictor variable instead of ‘family history of diabetes’. This way, we could use the same predictor for all three datasets. A disadvantage of this choice was that we could not use information on the diabetes status of siblings, which may have led to lower model performance. Next, in the dataset that was used to externally validate the model, there was no separate variable for ‘parental history of diabetes’. Instead, there was one combined variable for ‘family history of diabetes’. To evaluate whether this difference in variable definition could affect the performance of the model in the external validation dataset, we compared the performance of two models: first, a model including all regression coefficients of the DIRECT-DETECT prediction model, which included applying the ‘parental history’ coefficient to the ‘family history’ variable in the external validation dataset. In the second model, we applied all coefficients except the one for ‘parental history’ and allowed the model to freely estimate the coefficient for ‘family history’. We found that the performance of these two models in the external validation dataset did not considerably differ, indicating that the difference in variable definition had no notable influence on the results of the external validation. Thus, when pooling individual patient data, sensitivity analyses may be needed to evaluate the impact on model performance of differences in variable definitions between cohorts.

Confounding in etiological models and prediction models

A limitation of combining data from different populations is that differences between the populations may confound the observed relations [32]. Therefore, we addressed a possible confounding effect of different source populations by correcting the analysis for recruitment center in Ch. 8. In addition, the analysis was corrected for other confounders, which is common practice in etiological research [33]. Also in chapter 6 and 7, the etiological relations of interest were corrected for relevant confounders.

A methodological issue that we encountered is related to the role of confounding in the development of a prediction model. Although confounding is normally not an issue in prediction modeling –as observed associations between predictors and outcomes do not need to reflect causal relationships– we did observe a ‘confounding effect’ of cohort source when we combined data from three cohorts to develop the DIRECT-DETECT prediction model (Ch. 2). More specifically, when we developed the prediction model without correction for cohort source, a higher age seemed –unexpectedly– to be associated with lower HbA1c levels at follow-up, while after inclusion of cohort source in the model, this association was reversed. This could be explained by the fact that in one of the cohorts, compared to the other two cohorts, HbA1c levels were on average somewhat higher, while the mean age in that cohort was lower. These differences in HbA1c levels between the cohorts were possibly due to the different assays that were used between
the cohort studies to measure HbA1c levels. Using the prediction model and not correct for the differences in age and in HbA1c levels between the cohorts in this combined dataset would result in a model that would be less generalizable to other populations. We therefore concluded that it was necessary to include the cohort source variable in the model. This correction made it possible to accurately estimate predictor effects in a combined dataset of different populations. Overall, these results show that confounding is not only relevant in etiological models, but can also play a role in prediction models. Specifically, prediction models may not always be free of confounding when they are developed after data of different studies are combined.

Another issue related to confounding was encountered in chapter 4. Instead of evaluating whether the previously developed prediction model for residents treated with antibiotics could also be applied to untreated residents, we had considered to combine treated and untreated residents into one prediction model, including treatment with antibiotics as one of the predictors of short-term mortality. However, as we used data from observational studies, the decision to either treat a resident with antibiotics or to withhold treatment was not based on randomised study designs. In contrast, it has been shown that physicians might base this decision on illness severity [34]. Combining data from treated and untreated residents would therefore lead to confounding by indication. In non-randomized etiological research, such confounding by indication is often addressed by using propensity scores [35]. Propensity scores correct the analysis for the probability to receive treatment conditional on pre-treatment covariates [36]. For prediction modelling, it has been argued that there is no theoretical rationale to use propensity scores: confounding by indication might distort the estimations of the individual predictors, but the aim of a prediction model is to produce accurate predictions, not to accurately estimate predictor effects [37]. Confounding by indication is not expected to reduce the overall predictive ability of a prediction model. Indeed, it has been shown that propensity scores did not improve the predictive performance of three prediction models, compared to including all covariates in the model [37]. In this perspective, combining residents treated with antibiotics and residents not treated with antibiotics in one prediction model would not have been a problem for the predictive performance of the resulting prediction model. However, when such a prediction model would be used to decide whether or not to start treatment, rather than using the model after deciding whether or not to treat, to predict the risk of mortality, a non-accurate estimation of the treatment predictor could lead to incorrect conclusions about the effect of antibiotic treatment. Hence, for a prediction model that might be used to inform decisions on treatment, an accurate estimation of the treatment effect is important, and correction for confounding by indication would be necessary when combining treated and untreated residents in one model. Still, propensity scores can only correct for known and measured determinants of treatment [36]. In addition, a condition for obtaining unbiased effects by using propensity scores, is that for each individual, both treatment and non-treatment must be an option.
In other words, there must be an overlap in resident characteristics between the treated and the untreated group [36]. In our study, prior research had shown large differences in characteristics between treated and untreated residents [34, 38]. Therefore, we decided not to combine data of treated and untreated residents in one prediction model, to avoid erroneous interpretation of a –potentially biased– estimate of the treatment predictor as the effect of antibiotic treatment.

**Using routinely collected data for research purposes**

Next to cohort studies that were set up for research purposes, this thesis also includes two cohorts that consist of data that were –at least partly– collected as part of routine clinical practice. There are several advantages of using data that were routinely collected in clinical practice: this type of data is getting more readily available, represents a real-life care setting, and using these data for research purposes is cheaper than needing to set up a cohort [39]. A challenge that can be encountered when using routinely collected patient data is that, compared to performing measurements in a research setting, in clinical practice it is much more difficult to perform measurements in a standardized and objective way. Regarding our study on the association between self-reported hypoglycaemia and mortality (Ch. 7), some may argue that we should have measured hypoglycaemia objectively instead of using self-reported hypoglycaemia. Most research into the prevalence of hypoglycaemia has indeed been based on objective blood glucose measurements [15-17, 19-22], or only events were included that were registered in medical records [19-22]. However, for technical and practical reasons, measurements of glucose levels were not available from the patient records of the Hoorn Diabetes Care System Cohort. More importantly, in clinical practice, not all people with type 2 diabetes might regularly perform self-testing of their glucose levels. In practice, people might consult their general practitioner about experiencing hypoglycaemic sensations without confirmation by glucose measurement. Therefore, self-reported hypoglycaemia might better reflect hypoglycaemia as experienced in everyday life by people with type 2 diabetes compared to hypoglycaemia objectively measured in research settings. Similarly, in most nursing home cohorts that were included in this thesis (Ch. 4 and 5), the diagnosis of pneumonia was judged by the attending physician, mostly without evaluation of chest X-ray [14, 40-42]. Although some may argue that such an evaluation is necessary to confirm pneumonia, X-rays are not regularly performed in nursing home residents. Therefore, pneumonia as judged by the attending physician might better reflect pneumonia episodes as diagnosed in nursing homes.

A second challenge that can be encountered when using routinely collected patient data for research purposes, regards inconsistencies in coding [32]. For example, when using the data from the Diabetes Pearl cohort (Ch. 8), the distinction could not always be made between people with type 1 or type 2 diabetes. However, different mechanisms might play a role in different types of diabetes, and therefore, we wanted to exclude
people with type 1 diabetes from the study population. We have solved this by assuming a diagnosis of type 1 diabetes for people who were younger than 30 years at the time of diagnosis, and who were using insulin.

A third challenge when using routinely collected data is that data files might contain missing values [32]. In the Hoorn Diabetes Care System Cohort, we had to deal with missing data on hypoglycaemia, the determinant (Ch.7). For the main analysis, we decided that the first visit (out of a maximum of three available visits) with valid information on hypoglycaemia would be considered as the baseline visit. In other words, when information on hypoglycaemia was missing at the first available visit, the next visit was considered as the baseline visit. To evaluate whether this strategy had affected our results, we performed a sensitivity analysis where people with missing hypoglycaemia data on the first available visit were excluded from the analysis. We found that the results did not change when excluding these people and concluded that our strategy to consider the next visit as the baseline visit did not bias our results.

Finally, it should be noted that cleaning these data that were not collected for research purposes and checking the quality of the data was very time-consuming. It is very important to identify inconsistencies in coding and to evaluate whether missing values could be selectively missing. Moreover, the question always remains whether routinely collected patient data can be used to identify causal relations or whether there is residual confounding. We suggest that performing sensitivity analyses can help to understand and further explore observed associations in routinely collected patient data.

**Applying the same model in different populations**

After developing a prediction model, it is strongly recommended to perform an external validation study (i.e., applying the prediction model to data that were not used to develop the model), in order to assess the generalizability of the model [30] to a population that may differ from the development population in different aspects. Table 2 provides an overview of the external validation studies in this thesis, including the differences between the development cohort and the external validation cohorts, and the methods that were used to update the models in the different chapters.

In chapter 3 of this thesis, an existing prediction model that was developed in three Dutch cohort studies was externally validated in an Australian cohort study. The external validation cohort differed from the development cohorts in several aspects next to the geographical location (Australia vs the Netherlands). First, the external validation cohort was conducted more recently (2004–2012 vs 1989–2005), and drug treatment of risk factors had increased in this time [43, 44]. Second, the definition of cardiovascular disease was different, as no information was available on angina pectoris, peripheral arterial intervention, intermittent claudication or heart failure in the external validation cohort. Third, the follow-up period for non-fatal cardiovascular disease was shorter.
Table 2: Overview of external validation studies

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Outcome</th>
<th>Differences between development cohort and external validation cohort(s)</th>
<th>Updating</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Developing CVD, T2D or CKD</td>
<td>Geographical, time period, outcome definition, follow-up period, participant characteristics, outcome prevalence</td>
<td>Intercept</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Step 1: applying model for treated residents to untreated residents of the same cohort study</td>
<td>Intercept</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Treatment strategy, illness severity, resident characteristics, mortality rate</td>
<td>Intercept</td>
</tr>
<tr>
<td>4</td>
<td>Mortality in nursing home residents with pneumonia</td>
<td>Geographical, time period, resident characteristics</td>
<td>Intercept *</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Step 2a: applying updated model for untreated residents to untreated residents in three combined datasets</td>
<td>Intercept *</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Geographical, disease definition, resident characteristics, mortality rate</td>
<td>Intercept *</td>
</tr>
<tr>
<td>5</td>
<td>Mortality in nursing home residents with pneumonia</td>
<td>Time period, resident characteristics, mortality rate</td>
<td>Model revision + model extension</td>
</tr>
</tbody>
</table>

Abbreviations: CKD: chronic kidney disease; CVD: cardiovascular disease; T2D: type 2 diabetes

* Compared to step 1

These differences between the development cohorts and the external validation cohort might have contributed to the observed difference in participant characteristics and in prevalence of the outcomes. Also in chapter 4, there were several differences between the external validation populations and the development population. In this case, the nursing home residents in the development population and the first external validation population belonged to the same cohort study, but the prediction model was developed for residents with dementia and pneumonia treated with antibiotics, and was externally validated in residents not treated with antibiotics (step 1). The untreated residents were more severely ill than the treated residents, and they had a higher mortality rate. Next, four other cohort studies were used to externally validate the updated model for untreated residents. Two of these cohort studies differed from the development cohort geographically (USA vs the Netherlands), and three studies were conducted in a time period (2006–2007, 2004–2008, and 2007–2010 vs 1996–1998). In addition, diagnosis of pneumonia differed in one of the external validation studies: in this study, pneumonia was assessed by project nurses using clinical criteria, while in the other studies, pneumonia was diagnosed by a physician. In this same study, the outcome prevalence differed considerably from the other studies. Therefore, the model for untreated residents was validated separately in this cohort study (step 2b), while data from the other three cohort studies were combined into one external validation study (step 2a). Finally, in chapter 5, the same prediction model that
was developed for treated residents, was externally validated in a more recent cohort (2012–2015 vs 1996–1998). As mentioned earlier, care for older people changed since the development of the model, as well as population characteristics of nursing home residents. For example: the prevalence of malnutrition decreased, cardiovascular risk management improved, and new policies promote people to be cared for at home as long as possible. Indeed, resident characteristics differed between the development cohort and the external validation cohort, as well as mortality rates.

In chapter 3 and 4, we observed that the discriminative ability of the models was acceptable in the external validation populations, indicating that the models could adequately discriminate between high risk and low risk people in populations that differed from the development population in participant characteristics, prevalence of the outcome, and even in treatment strategy. Thus, despite differences between the development population and the external validation populations, the same predictors could adequately explain differences in outcome risk between individuals within the development population or within the external validation population. However, the calibration of the models (i.e., their ability to accurately predict absolute risks to develop the outcome) was inadequate in the new populations with different outcome rates compared to the development population, resulting in systematic overestimation or underestimation of outcome rates. Therefore, there was a need to update the models to improve their performance in the new populations. Steyerberg has described different methods to update prediction models, including methods for recalibration, model revision and model extension [45]. Methods for recalibration include intercept adjustment with or without adjustment of the slope and can be applied to improve the calibration of a model in an external validation population. Methods for model revision and/or model extension may improve the discriminative ability of a model, and include re-estimation of one or more predictors in the original prediction model with or without extension of the model with one or more additional predictors [45]. In chapter 3 and 4, we applied methods of recalibration to improve the calibration of the models [45]. In line with previous research, we observed that the simplest method for recalibration, namely adjusting the intercept of the models, was sufficient to improve calibration in the new populations [45-47]. The fact that discrimination was adequate in the external validation populations while calibration was inadequate, indicates that the models’ predictors could adequately explain differences in outcome risk between individuals within the same population, but could not explain differences in outcome risk across populations. Theoretically, this indicates that new predictors might be identified that can explain differences between populations. However, it is also possible that differences in outcome rates exist between the populations as a result of changes in treatment. For example in chapter 3, the external validation cohort was more recent, while the use of antihypertensives and statins has increased over time [43, 44], which likely contributed to the lower prevalence of chronic cardiometabolic diseases in the external validation cohort. In practice, this means that in
order to accurately use the same prediction model in different populations, the model intercept should be recalibrated when the model is applied to a population that is expected to differ from the development population in characteristics (also referred to as case-mix [45]), or in prevalence of the outcome.

In chapter 5, we observed that the performance of the model was acceptable in the more recent validation cohort, indicating that the prediction model was rather robust for changes over time in care practices and resident characteristics. However, the discriminative performance was slightly lower than in the development cohort, which might be explained by changes in the care for older people and in resident characteristics in the external validation cohort compared to the development cohort. The lower discriminative performance in the more recent cohort indicates that new predictors might help explain the differences in mortality risk or that predictor estimates for the current predictors may need to be revised for the external validation population. Thus, we evaluated the performance of the model after more extensive updating of the model, by using methods for model revision and model extension [45]. Removing one predictor from the model while extending the model with four new predictors improved the discriminative ability in the more recent cohort. These results show that even after 20 years, with changes over time in care practices, a prediction model might still perform adequately, although discrimination in a more recent cohort might be improved by model revision and model extension.

To compare models before and after an update, reclassification tables were used in chapter 4 and 5. These tables show the number of people that are classified as ‘low risk’ or ‘high risk’ and that shift (or: are reclassified) from low to high risk (and vice versa) by using a new model compared to an old model [48]. The net reclassification index (NRI) can be used to quantify the extent of reclassification. A disadvantage of reclassification tables and the NRI is that an a-priori cut-off point has to be chosen to define low risk versus high risk, and that different cut-off points may lead to large differences in the results [30, 49]. The use of a continuous NRI has been proposed to overcome this problem [50]. This measure does not rely on categories and considers any shift in predicted risk for each individual. However, most changes in predicted risk might not result in changes in treatment in clinical practice, and therefore, concerns have been raised that this measure might be difficult to interpret in relation to clinical practice [30, 49, 51]. In spite of its limitations, NRI can be an informative clinical performance measure when it is guided by a clinically relevant cut-off point [30, 49]. In chapter 4, we were able to use such a cut-off point for the predicted 14-day mortality which Dutch elderly care physicians considered to be clinically relevant, namely 80% [52]. In chapter 5, we could not use the same cut-off point, as in this less severely ill population, no predictions were made >80%. We have therefore chosen a lower cut-off point, namely 50%.

A final issue regarding the application of the same prediction model to different populations, is that one may wonder why we did not develop a new model for patients treated differently in Ch. 4. Indeed, it seems to be common practice to develop new models
for each population or subpopulation, with over 60 models to predict breast cancer prognosis [53], and over 40 models to predict incident type 2 diabetes [54]. However, having to choose between so many competing prediction models is very impractical for clinicians. Moreover, the fact that a simple method for recalibration might suffice to make an existing prediction model applicable to patients treated differently, implies that there is no need to develop new prediction models for each new population. This is in line with the TRIPOD statement, which states that existing prediction models should be validated instead of developing new models for each population, this way combining prior knowledge with new knowledge [30].

**IMPLICATIONS FOR CLINICAL PRACTICE AND FOR RESEARCH**

The results described in this thesis have implications for clinical practice and raise questions for further research. In addition, the lessons that we learned regarding the methodological issues that we encountered during the research described in this thesis may help researchers to think about how to apply these methods in their research, and may thus have implications for research practice.

The model for predicting chronic cardiometabolic disease risk (Ch. 3) showed good model performance in the development study as well as in our external validation study. Therefore, this model can be used as a first step in screening to identify individuals who are at increased risk for the development of chronic cardiometabolic diseases. The model is freely available for patients via health organization websites, referring these high-risk individuals to their general practitioner, where they can further be screened for risk factors, including standard blood tests. In addition, the model is incorporated into the Dutch guidelines for general practitioners, ‘The Prevention Visit’ [55]. As a next research step, the effectiveness in terms of number of newly detected patients with chronic cardiometabolic disease and changes in individual risk factors, as well as the cost-effectiveness of using this prediction model in combination with a tailored lifestyle intervention are currently being studied [56, 57]. For the models predicting short-term mortality for nursing home residents with dementia and pneumonia (Ch. 4 and 5), further research may focus on studying the usefulness in clinical practice of the two prediction models for residents treated with antibiotics and those untreated. Do physicians find it useful to be informed on prognosis by these prediction models, do they use the models to inform residents and their families on prognosis, and do the models support decision making?

We found that several prediction models did not accurately predict absolute risks in external validation studies with different outcome rates compared to the development population. Recalibration of the model was needed to improve absolute risk predictions. These findings lead to important considerations for clinicians. When clinicians want to use a prediction model, they should consider whether the population in their general
practice is expected to be comparable to the population that was used to develop and/or externally validate the prediction model. For example, can the population characteristics, such as characteristics related to illness severity, be expected to be comparable? Has the treatment of risk factors changed substantially since the development of the model, or is there a considerable difference in treatment between the country of this clinician and the country where the model was developed? And overall, is the prevalence of the outcome of interest expected to be comparable between the patient population in the general practice and the population that was used to develop and/or externally validate the prediction model? If considerable differences are expected, the prediction model should first be calibrated for the new population before using the model in this population.

Finally, our experiences regarding the use of data that were collected as part of routine clinical practice may help researchers to consider how to use this type of data for research purposes. Routinely collected patient data are increasingly being used for research purposes, as these data are getting more readily available and it is a relatively cheap way to capture large amounts of data, including many clinical events in large populations [39, 58]. Nevertheless, we found that it was very important to critically look at routinely collected patient data. Further, checking the quality of the data and cleaning the data were very time-consuming. For example, there were missing values in data files and we had to evaluate whether these values could be selectively missing. Also, we identified coding inconsistencies and had to find ways to deal with these inconsistencies. In addition, as in observational data from other sources, confounding plays a large role in routinely collected patient data, and thus, conclusions about causality cannot be inferred from this type of data. Researchers should be aware of these challenges when using routinely collected patient data for research purposes. We found that performing several sensitivity analyses helped to better understand the impact of inconsistencies in coding and missing data on the results, as well as the impact of possible confounding factors. Although the question remains whether associations identified using routinely collected patient data reflect causal relations or whether there still is residual confounding, we would suggest that routinely collected data can be used to generate hypotheses that can be tested further in studies using more controlled designs.
CONCLUSIONS

Predicting type 2 diabetes using only non-invasive predictors
• The model for predicting absolute HbA1c levels showed low explained variance and discrimination, limiting its use as a screening tool in clinical practice.
• The model for predicting chronic cardiometabolic disease risk showed good model performance in the development study, as well as in our external validation study.
• This model can be a useful tool in large-scale prevention programs, as a first step in screening to identify individuals who are at increased risk for the development of chronic cardiometabolic diseases. In a second step, these high-risk individuals should further be screened for risk factors, including standard blood tests.

Predicting mortality in nursing home residents with dementia and pneumonia
• After a simple update of the model intercept, the same predictors that were relevant to predict short-term mortality for nursing home residents with dementia and pneumonia treated with antibiotics, could also adequately predict short-term mortality for untreated residents, despite large differences in characteristics between treated and untreated residents.
• The same model for treated residents could still validly be used after 20 years. An extended model performed better in a more recent dataset of treated residents, compared to the original model.
• The two models –one for treated and one for untreated residents– can be used in clinical practice to inform physicians, residents and their families on prognosis and may this way support decision-making.

Etiological models in the field of type 2 diabetes
• In younger people, larger social jetlag was associated with a higher prevalence of cardiometabolic risk factors and diabetes/prediabetes. In older people, no significant associations were found.
• People with type 2 diabetes who reported hypoglycaemia not requiring medical assistance did not have an increased risk of short-term mortality, compared to people who reported no hypoglycaemia. Further research might focus on the potential role of hypo-unawareness in this association.
• Partner's socio-economic status was associated with HbA1c levels in people with type 2 diabetes, suggesting that physicians or nurse practitioners should not only consider the socio-economic status of a person with type 2 diabetes when assessing the risk of poor glycaemic control, but also the socio-economic status of the partner.
Methodological issues

• When combining individual patient data from different cohort studies, sensitivity analyses are needed to evaluate the impact of differences in variable definitions between cohorts.

• Confounding does not only play a role in etiological models, but also when developing prediction models in a combined dataset of different cohorts, even though causality is not the focus in prediction modelling. Confounding by cohort source should be considered and evaluated when a prediction model is developed in a combined dataset of different cohorts.

• Using routinely collected data for research purposes might better reflect the clinical care setting compared to data from an experimental research setting. Cleaning this type of data and checking the quality of the data are very time-consuming. Sensitivity analyses can help to understand and further explore observed associations and the role of (residual) confounding.

• A simple method for recalibration sufficed to make two existing prediction models applicable to different populations, even when the external validation population was treated differently. In a third study, a model still performed adequately in a more recent population, and performed even better after more extensive updating of the model by revising existing predictors and extending the model by adding extra predictors. This is in line with previous recommendations that existing prediction models should be validated and—if necessary—updated, instead of developing new models for each population.

• When using a prediction model in clinical practice, it is important to evaluate whether patient characteristics and outcome prevalence in this population are expected to be comparable to the population that was used to develop or validate the model. The model intercept should be recalibrated before applying a model to a population that is expected to differ from the development population in characteristics or in prevalence of the outcome.

Overall conclusions

• Several research questions that are relevant for the field of type 2 diabetes and the field of pneumonia in nursing home residents with dementia were answered in this thesis using prediction models and etiological models.

• Next to answering research questions that are relevant for specific disease fields, we found that focusing on the application of existing methods and focusing on methodological issues that are encountered when applying these methods, can have an added value for future research in comparable data situations.

• Researchers and clinicians should consider these topics when applying and interpreting prediction models and etiological models to their specific disease fields.
REFERENCES


14. van der Steen JT, et al., *Predictors of mortality for lower respiratory infections in nursing home residents with dementia were validated transnationally*. Journal of Clinical Epidemiology, 2006; 59(9):970-79.


Chapter 9

34. van der Steen JT, Ooms ME, Adèr HJ, Ribbe MW, and van der Wal G, *Withholding antibiotic treatment in pneumonia patients with dementia: A quantitative observational study*. Archives of Internal Medicine, 2002; 162(15):1753-60.


