Hacking stroke in women: towards aetiology-driven precision prevention

Os, H.J.A. van

16. General discussion

In **part I** of this thesis, I conducted aetiological research on sex differences in the pathophysiology and clinical presentation of stroke. In **part II**, the aim was to develop female-specific prediction models of the risk of stroke that can be implemented in clinical workflows.{Steyerberg, 2014 #2435} In the current chapter, I will discuss the overarching challenges addressed in part I and part II. Because the overarching goal of this thesis is to lay the foundation for precision prevention of stroke in women, this chapter seeks to show what future research and policy steps are needed to move from this foundation to large-scale implementation of precision prevention in clinical practice.

Pathophysiology of sex differences in stroke – what should the next steps be?

The modulating effect of age on the risk of stroke

Age is the most important predictor for the risk of cardiovascular events, and substantially modulates the relationship between other traditional cardiovascular factors and cardiovascular risk.[2,3] Much is still unknown about whether – and if so, how – age modulates the effect of female-specific risk factors on the risk of stroke. In women with migraine aged 45–49 years, the risk of ischaemic stroke is substantial (aHR: 5.35, 95% CI: 2.08–13.79) compared to women without migraine, while there appears to be no increased risk in women with migraine older than 65.[4] Also in women with a history of preeclampsia, the increased risk of stroke is evident during reproductive age, although some studies suggest that the risk increase is also present later in life.[5,6]

High-quality data on the associations between female-specific risk factors and stroke in women under 50 years are scarce, which is a recurrent limitation in part I of this thesis. In **chapter 2,** we used DUST study data to evaluate the hypothesis that the association between migraine and stroke is caused by an increased burden of cerebrovascular atherosclerosis in people younger than 50 years. However, this was complicated by the fact that only 86/656 (13%) patients were younger than 50 years. We did, however, see a trend of fewer cerebrovascular calcifications in patients with versus without migraine (aRR: 0.82; 95% CI: 0.64–1.05), which decreases the likelihood that macrovascular atherosclerosis is the main pathophysiology underlying the migraine–stroke relationship. In **chapter 5**, I studied the potential association between migraine and delayed cerebral ischaemia in patients with aSAH. Although in the overall population no association was present, I found a statistically significant interaction ($p = 0.075$, at an alpha of 0.10 for interaction terms) between age and migraine with respect to risk of delayed cerebral ischaemia. This suggested an association between migraine and delayed cerebral ischaemia exclusively in young patients. However, after including additional aSAH

patients in the cohort to perform an analysis in patients under 50 years **(chapter 6)**, we also did not find an association between migraine and the risk of delayed cerebral ischaemia. This contradicts an earlier matched case-control study, which found an odds ratio of 2.68 (95% CI: 0.99–7.29; p-value < 0.05) for the association between delayed cerebral ischaemia and a history of migraine in young women.[7] The point estimate from this case-control study could be an overestimation because it was derived from a small dataset of matched pairs, further illustrating the importance of replicating observational association studies with cohorts of sufficient sample size.[8] In the String-of-Pearls Stroke Study cohort presented in **chapter 9**, only 65 (7%) of the 997 included women with a history of migraine were younger than 50 years. This sample size did not allow for the assessment of multiple associations between traditional cardiovascular risk factors or ischaemic stroke aetiology and migraine. I did, however, demonstrate that female stroke patients with migraine were at increased risk of stroke onset at an age under 50 years (RR: 1.7; 95% CI: 1.3–2.3). Since no sex differences in the occurrence of traditional cardiovascular risk factors were found in women with- versus without migraine, this suggests that mechanisms other than traditional macrovascular atherosclerotic aetiologies play a role in the higher risk of stroke in migraine.[4]

Large, population-based cohorts may help solve questions regarding stroke pathophysiology in women under 50 years. In **chapter 7,** I assessed the potential interaction between combined oral contraceptive use and migraine leading to ischaemic stroke in women aged 18–50 years. In the large EHR-derived population-based STIZON cohort (n = 1,404,681 adult patients), a total of 617 women under 50 years at baseline suffered from stroke during follow-up, which resulted in a sufficient sample size to answer my research question. In **chapter 14**, the STIZON cohort enabled me to assess associations between several female-specific risk factors and ischaemic stroke risk factors in three subcohorts of women aged 20–29, 30–39, and 40–49 years. I found that migraine (aHR: 1.48; 95% CI: 1.1–2.01) and preeclampsia (aHR: 1.83; 95% CI: 1.17–2.85) were only significantly associated with risk of stroke in women aged 30–39 at baseline, while hormonal contraceptive use was associated in both women aged 30–39 (aHR: 1.20; 95% CI: 1.03–1.40) and aged 40–49 (aHR: 1.35; 95% CI: 1.22–1.49). These findings corroborate the hypothesis that the causal relationship between female-specific risk factors and ischaemic stroke is substantially modulated by age.

Data quality regarding exposure to female-specific risk factors – migraine as an example

For aetiological studies on the migraine–stroke association, another frequent limitation is the quality of the data on migraine. Migraine is a clinical diagnosis, and the gold standard is an interview performed by a headache specialist, based on the

International Classification of Headache Disorders criteria.[9] However, such interviews are not structurally performed in routine care, and are not feasible in an acute stroke research setting. Therefore, in **chapters 2, 3, 5, 6, and 9,** history of migraine was assessed using the previously validated five-question Migraine Screener for Stroke (MISS). The MISS has a high positive predictive value for migraine with and without aura combined (0.80; 95% CI: 0.59–0.93), but a poor positive predictive value for migraine with aura (0.38; 95% CI: 0.24–0.53).[10] Consequently, I could not sufficiently discriminate between migraine with and without aura in the aforementioned five studies in **part I** of this thesis. This distinction is important because a systematic review of cohort studies found that migraineurs with aura (aHR 1.56, 95% CI: 1.30–1.87) but not those without aura (aHR 1.11, 95% CI: 0.94–1.31) had an increased risk of stroke.[11] In contrast, one large Danish cohort study that was published later showed a significant association between migraine without aura and ischaemic stroke. However, this study only included migraine patients who were diagnosed in an emergency department, inpatient, or outpatient hospital setting, which may have resulted in a selection of migraine patients that suffer from relatively frequent migraine attacks. A high attack frequency has been associated with a relatively stronger increase in the risk of ischaemic stroke.[12] Further, in our studies using the MISS we cannot exclude the possibility that it is mainly or exclusively the subset of migraine patients with aura that is responsible for the associations found between migraine and radiological characteristics of stroke, or of stroke itself. This could result in an overestimation of the risk of stroke in patients with migraine.

In **chapters 7, 14 and 15**, migraine history was defined based on routine care data, using a combination of primary care and hospital diagnosis codes and medication prescription data for specific antimigraine drugs. **Chapter 14** shows that in the population-based STIZON cohort of women under 50 years (n = 409,026) migraine according to my definition was present in 4% of individuals. This is much lower compared with the results from large population-based migraine studies, which report a one-year prevalence of migraine of around 18–25% in women and 6–8% in men. [13-16] The underreporting of migraine prevalence in routine care data sources may have different causes. One study from 2002 estimated that only two-thirds of migraine patients visit a physician for their headache.[17] If a patient visits the GP and migraine is diagnosed, there still is a substantial probability that the diagnosis is not registered in primary care, since underreporting in the EHR is generally high.[18] A less error-prone source of routine data in the Netherlands is the out-patient pharmacy prescription database which is linked with the STIZON primary care and hospital data.[19] By using medication prescription codes, it is possible to accurately identify the subgroup of migraine patients who use migraine-specific drugs. This subgroup of migraine patients likely has more severe migraine attacks. However, no

conclusive evidence exists that the severity of migraine attacks modulates the risk of ischaemic stroke.[20] Another important drawback of the EHR as data source for definition of migraine is that the distinction between migraine with and without aura is rarely made in clinical practice.

The bias-variance trade-off, and the role of the EHR in aetiological research

For unbiased estimates of associations between female-specific risk factors and stroke, the collection of high quality data on both exposure and outcome is essential. Simultaneously, a sufficient sample size is required, which is challenging because stroke in women under 50 years is rare. In practice, however, the larger the sample size, the poorer the quality of exposure and outcome assessment, generally due to the practical and financial constraints of large-scale data collection. Therefore, most studies – including those in this thesis – suffer from a data quality-quantity trade-off. Statistically speaking, this is the trade-off between bias and variance, and this trade-off is one of the most important factors in choosing the right research design.[21] Because stroke in women under 50 years is rare, a case-control design is relatively efficient. If controls are sampled in an optimal way, a case-control study should result in the same estimation of an exposure–outcome relationship as a cohort study.[22] However, in practice multiple co-occurring sources of bias often limit causal inference from case-control data. A prospective cohort study design could eliminate problems such as recall bias. An example of a traditional prospective cohort study in which the association between female-specific risk factors and stroke is assessed, is the Nurses' Health Study II (115,541 women aged 25–42 years at baseline). However, these large-scale studies are logistically challenging, and the assessments of exposures are limited. In the Nurses' Health Study II, information on migraine aura or migraine frequency were not available, which eliminated the possibility for clinically relevant subanalyses.[23]

In the search for very large, richly phenotyped cohorts, pooled routine health data collections may offer a part of the solution, at a relatively low cost. Therefore, we use the STIZON cohort in **chapters 7 and 13–15**, which has the important advantage that it consists of multiple linked routine data sources including the primary care EHR, hospital ICD-9 and ICD-10 registrations, and outpatient pharmacy data.[90] For female-specific risk factors, the primary care EHR is the primary source of interest. However, a major drawback is that the data entry into the EHR has almost no quality assurance measures, which results in lower quality of outcome and exposure data.[18,24,25] Underreporting of exposures in the EHR is a common problem, which may result in biased estimates of outcome-exposure relationships. At the same time, the confidence interval around such biased estimates may be very small because of the large sample size. It is therefore important to not fall victim to the so-called big data paradox, which has been described as 'the more

the data, the surer we fool ourselves'.[21] Which directions should we take to answer the most urgent aetiological questions on the associations between female-specific risk factors and stroke in women? And which (observational) research strategy offers the most optimal trade-off between bias and variance? There lies a complementary value in both (i) EHR-based cohorts with large sample sizes and ascertainment of a very wide range of exposures but low data quality; and (ii) traditional cohort or case-control studies with a higher quality of ascertainment of outcome and exposures but often with limited sample sizes, depending on the aetiological research question. Therefore, there is a role for EHR-based cohort studies in aetiological research.

Prediction models for the risk of stroke in women – external generalisability, transportability, and the role of causality

The bias-variance trade-off in prediction research

In **chapters 12, 14 and 15**, I used EHR-based cohorts to develop prediction models, for the estimation the risk of cardiovascular events and stroke specifically. In prediction research, the bias-variance trade-off also plays a key role in research methodology, although the concept is applied in a fundamentally different way compared with aetiological research. Bias in prediction does not mean the unbiased estimation of exposure–outcome associations, but constitutes a difference between predicted and observed risks in a target population of interest. Important flaws in EHR-based cohorts – such as measurement error in exposures – may lead to a biased estimation of an exposure–outcome relationship, but do not necessarily lead to biased predictions. That is, if the measurement error of the derivation population is similar to that of the target population.[26] Bias in predicted risks result from limited generalisability or transportability of a prediction model to a new setting, and restricts the clinical usefulness and -safety of a prediction model. Therefore, the careful reporting of discrimination and calibration after internal and external validation is key. The calibration of the model is particularly important for clinical decision making, because a model needs to be well calibrated around the absolute predicted risk that constitutes the decision threshold.[27] In **chapter 12**, I assessed the impact of different choices in data preparation on model discrimination and calibration. This study essentially simtulates external validation scenarios in which a clinical prediction model is derived from EHR data, and then transported to another context in which different data preparation choices are made. The conclusion is that differences in data preparation can have a large impact on model calibration. However, model calibration is not consistently reported in clinical research practice. Of 363 cardiovascular risk models that have been identified up until 2018, for as few as 21% calibration is reported after external validation.[28]

Consequently, the lack of reporting of calibration has previously been described as the Achilles heel of current predictive analytics.[29]

Another decisive factor that relates to the bias-variance trade-off in prediction modelling is overfitting, also known as model optimism. Overfitting means that a model performs well on the study data but predictions are not valid for new subjects.[30] The extent of overfitting can be reduced by increasing the number of events per each predictor included in the prediction model. For linear regression models, a widely accepted rule of thumb is a minimum of 10 events per predictor.[31] However, for complex data-driven models, more than 200 events per variable may be needed to limit overfitting.[32] Because complex data-driven models can be notoriously 'data-hungry', it is important to reduce risk of overfitting as much as possible. Therefore, in **chapter 13,** I presented a modelling pipeline to limit the risk of overfitting. I did this by preventing so-called information leakage between the optimisation of hyperparameters (inner cross-validation loop) and internal validation of the optimised models (outer cross-validation loop).[33] I reused this code for **chapters 14 and 15,** and made it publicly available online (*https://github.com/L-Ramos/MrClean_Machine_Learning*).

The trade-off between model complexity and interpretability

Complex data-driven models have the advantage of capturing non-linear predictor-outcome relationships and high-dimensional interactions among predictors, which could lead to additional explained variance of the outcome but may also result in overfitting.[34] Recently, a systematic review of 71 studies that compared model performance between complex data-driven- and traditional linear regression models based on traditional research cohorts, showed that in the majority of studies there is no additional value of complex data-driven models.[35] In **chapter 13** I, too, did not find an added value of complex data-driven models compared with logistic regression for the prediction of functional outcome in patients after endovascular treatment for an ischaemic stroke. Interestingly, another study which compared complex data-driven models with traditional regression models for prediction of the same outcome in a similar population did find a significantly higher performance for complex data-driven models (C-statistic: 0.86; 95% CI: 0.85–0.86) compared with traditional statistical methods (C-Statistic: 0.79; 95% CI: 0.77–0.81). Of note, in this study multimodal imaging predictors such as CT perfusion parameters were included in complex data-driven models, which may have resulted in a more complex data environment compared with the study in **chapter 13.**[36] Is this difference in discriminatory performance (C-statistifc of 0.86 versus 0.79) clinically relevant, and does it justify the increase in model complexity? To answer this question, discrimination is not the most relevant performance indicator that should be used. Rather, difference in clinical utility of two models needs to be compared,

for example by showing reclassification around the absolute risk that constitutes a relevant threshold for a decision in clinical practice.[27] In **chapter 15**, complex data-driven methods for predictor selection and for prediction of first-ever cardiovascular events in the general population under 50 years were compared with a Cox PH reference model including only traditional cardiovascular predictors. The best performing model was a Cox PH model including most important 50 predictors resulting from a Cox ElasticNet predictor selection procedure, which resulted in a correct reclassification of 3.7% of events around a relevant decision threshold (2.5% absolute 10-year risk of cardiovascular events) compared with the reference model. The benefits of the large scale implementation of this model may justify the increase in model complexity compared with the reference model. Importantly, **chapter 15** shows that the added value of complex data-driven may not arise from the increased flexibility of the model, but of selection of predictors other than those based on prior knowledge.

Importance of aetiological knowledge to support model transportability

The aim of statistical modelling is traditionally distinguished in (i) explanatory analysis, i.e. assessing causal relationships between exposures and outcomes, and (ii) predictive analysis, which aims to accurately predict individual risk on an outcome based on input data.[37] All causal factors are per definition predictors of the outcome, but predictors are not necessarily causally related to the outcome and can also be a proxy for causal factors.[38] Having yellow fingers is an example of such a predictor; highly predictive for the risk of cardiovascular events, but its association with the outcome is confounded by smoking.[39] Causal knowledge of predictor–outcome associations may be important to ensure model transportability. Transportability requires that the prevalence of the outcome, distribution of the predictors and predictor–outcome associations must be conserved.[1] Not meeting these requirements leads to a deterioration in model calibration, called calibration drift.[40,41] Therefore, for the development of prediction models it is preferred to choose predictors that pose a limited risk of calibration drift. In **chapter 15**, I included binary missing indicators of the presence or absence of a vital parameter or laboratory measurement, such as systolic blood pressure or total serum cholesterol. I did this because the measurements themselves are only present in a small proportion of the general population. Because in EHR data the missingness mechanism is often missing not at random, and the imputation of predictor values may lead to biased predictions.[42] The presence of such measurements are a result of GP behaviour and can be informative. However, to use a missing indicator for presence or absence to improve the predictive performance of models can be dangerous, because a change in GP behaviour may affect the predictor–outcome association over time, causing calibration drift. [43,44] An example is the 2019 update of the cardiovascular risk management guideline from the Dutch College of General

Practitioners. In the update it was recommended to start the screening of high blood pressure in women ≥ 45 years with a history of preeclampsia or gestational hypertension.[45] As a result, GP behaviour will likely have changed over the last years, together with the association between presence of a blood pressure measurement and cardiovascular outcomes. A model developed before the update, but applied to patients afterwards, may have suffered from calibration drift. To prevent calibration drift, we should quantify model performance over time, and may need to periodically update prediction models that have been implemented in clinical practice.[46]

Knowledge discovery through data-driven analysis

Identification of novel, strong predictors, or interaction among different predictors of the risk of cardiovascular events may lead to new directions for aetiological research – so-called knowledge discovery. Complex data-driven models can be used to identify strong predictor-outcome associations, and could be the first step to develop causal hypotheses. For example in **chapter 15,** I used sex-specific, EHR-derived Cox elastic net models to rank predictors according to their importance. In this ranking, substantial differences between men and women could be seen, including two female-specific risk factors which were ranked in the top 20 (combined oral contraceptive and intrauterine contraceptive use). Interestingly in men, psoriasis was ranked among the most important predictors. Previous studies indicate a potential causal relationship between psoriasis and cardiovascular disease, potentially because of a systemic inflammatory component of the condition.[47] Epilepsy is another non-traditional cardiovascular predictor that is ranked within the top 20 of most important predictors for men. However, the association between epilepsy and the risk of cardiovascular events may be confounded by health behaviours, which illustrates the need of replication of the exposure–outcome associations in aetiological research.[48]

Implications, and future research and policy steps

Do we need yet another cardiovascular risk prediction model?

A systematic review in 2016 identified 363 models that predict the risk of cardiovascular events, of which 70 (19%) have been externally validated. In general, it is advisable to reuse already validated models, or perhaps tailor them to local settings using recalibration.[28] However, in **chapters 14 and 15** I chose to build new EHR-based prediction models for the risk of cardiovascular events and stroke specifically, because of the following reasons.

First, the SCORE2 model for the estimation of cardiovascular risk in the general population does not include patients under 40 years, and no female-specific

predictors are used. Therefore, in **chapter 14** I derived three prediction models from populations of women aged 20–29, 30–39, and 40–49 years, which included several female-specific and psychosocial factors. These models may help select young women with an absolute risk of stroke that is relatively high compared with women of the same age that have ideal risk factor levels. To this purpose I developed a novel 'stroke risk age' tool. Currently, a lack of risk awareness represents a major restriction on the implementation of preventive measures and healthy lifestyle choices in the female population.[49] Although for adults under 50 absolute risks rarely warrant the prescription of preventive medication according to the ESC guideline[50], the stroke risk age tool may help to identify women who could benefit from lifestyle interventions that target modifiable risk factors.[50,51]

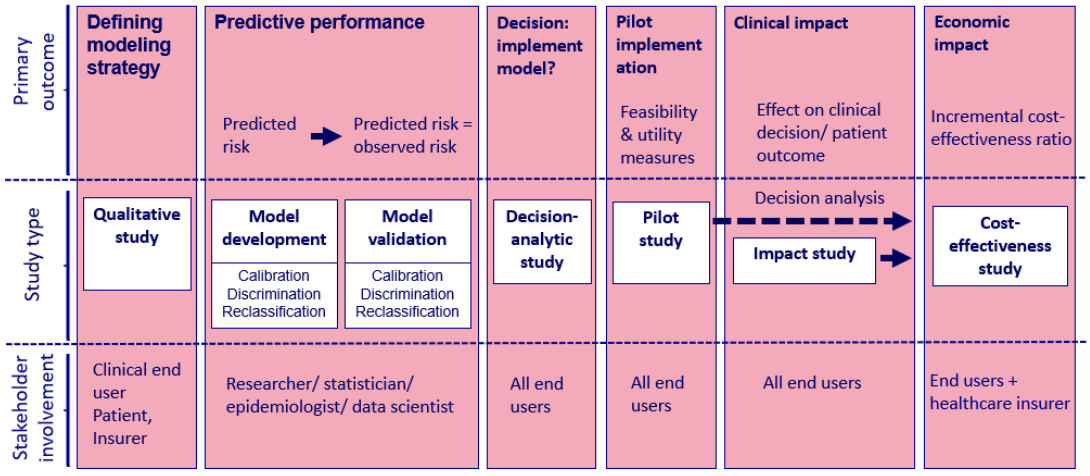
Second, we should distinguish between the use of a prediction model on the individual patient level and on the population level.[52] Regarding the individual patient level, predicted cardiovascular risk may for example be used to prescribe preventive statin therapy, in which case the cost of over- or underestimating risk for such an impactful clinical decisions is high.[18] Choosing a model that has been externally validated multiple times, such as the SCORE2 model, may therefore be the optimal choice.[53] In addition, treatment benefit based on the reduction of lifetime risk of cardiovascular events may be of added value, using for example the LIFE-CVD model.[57] The predictor values that are required as the input for SCORE2 or LIFE-CVD – such as systolic blood pressure or total serum cholesterol – can be collected directly from the patient during a consultation in the primary care practice centre. On the population level, prediction models can be used to generate a distribution of absolute cardiovascular risks of an entire primary care practice population. This enables the GP to proactively invite the patients with the highest risk, who may currently be undertreated.[54] However, for more than 80% of the Dutch adult primary care population no information on important cardiovascular predictors such as systolic blood pressure and total serum cholesterol is present in the EHR in a one-year time period. It is not feasible to invite all of these patients to the primary care practice centre for laboratory or vital parameter assessments. Imputing missing values in the EHR data is also problematic, as was discussed previously. Therefore in **chapter 15,** the sex-specific prediction models for the risk of cardiovascular events used predictive information from predictors other than continuous measurements of traditional cardiovascular predictors, which means there was no need to impute missing values. Instead of using continuous measurements such as systolic blood pressure as predictors, missing indicators were used. Therefore, despite inherent quality issues of EHR data, new EHR-based models can complement existing cardiovascular risk models that are based on traditional cohorts. Both EHR-derived and traditional cohort-derived models could be used in a two-step approach, at the population and individual level. First, the

primary care EHR is automatically screened for women with a potentially high risk of cardiovascular disease or stroke (population level) using models as presented in **chapter 15**. Second, the women identified in step one are invited for a cardiovascular risk (re)assessment at the primary care practice (individual level). During this patient visit, relevant predictor data, such as systolic blood pressure, can be collected to use as input for models that are extensively validated such as the SCORE2 or LIFE-CVD model.[53,55]

Towards the wide-scale implementation of clinical decision support systems

In **chapters 14 and 15**, the development and internal validation of several EHR-based models have been described. However, a long 'valley of death' lies ahead before these model could be implemented in clinical practice on a wide scale. Next steps include external validation,[1] and the assessment of the impact of the model implementation on health outcomes, for example via a clinical (cluster-)randomised trial.[52] Impact assessment is a complex and costly step. Therefore, a decision analytic study may be carried out as step in between model validation and implementation, to investigate the potential impact on health and economic outcomes by modelling the likely effect of its implementation on clinical decision making.[52] When one decides to perform an impact assessment of the model, the impact can only be validly estimated when the prediction model is successfully implemented in clinical practice, often in the form of a clinical decision support system (CDSS).[56] The implementation of a CDSS is a multifaceted challenge, which requires automated integration with the IT infrastructure and therefore clinician's workflow, interpretability of prediction model outputs by the end user, and more.[57]

Figure 4. The roadmap of a clinical prediction from model conception towards the wide-scale implementation in the clinical workflow as a CDSS product



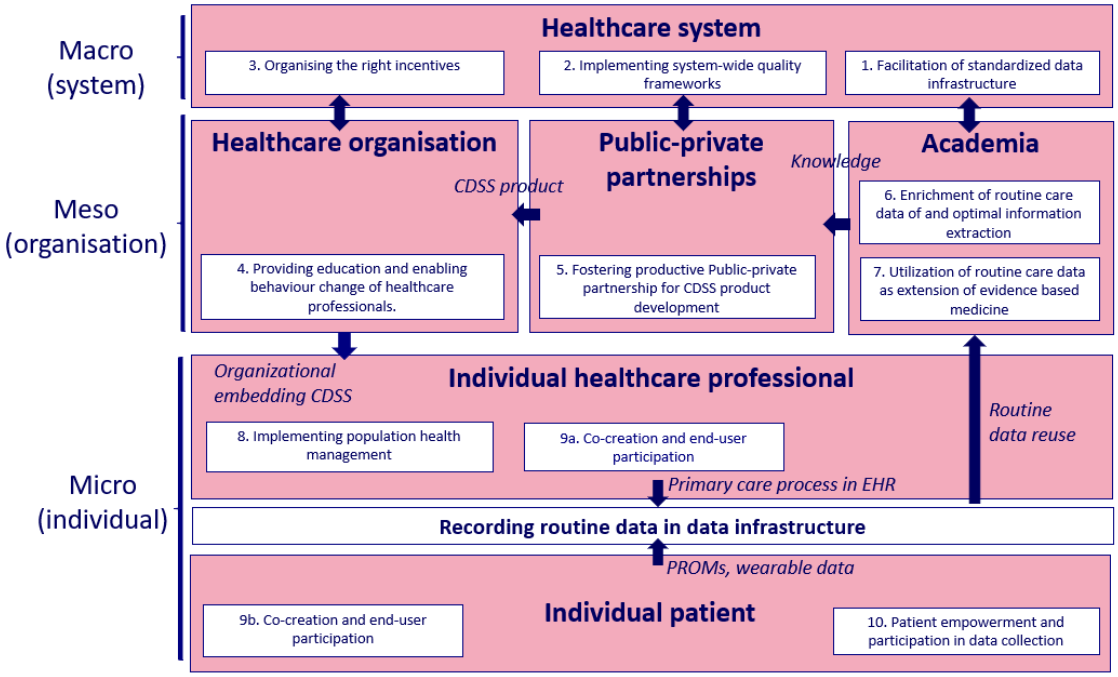
Precision prevention of stroke in women - towards a learning population health management system

The Lancet International Commission on Women and Cardiovascular Disease recently set the goal of significantly reducing the global burden of cardiovascular disease, including stroke, in women by 2030.[58] To improve the clinical outcome of stroke in women, stroke management including timely diagnostics should be improved, including an increase in awareness of differences in stroke symptoms as presented in **chapter 10**. But to make a substantial impact, the focus must shift from treatment to prevention of stroke, because approximately 70% of cardiovascular disease cases and cardiovascular disease-related deaths in the general population can be attributed to modifiable risk factors.[59] The lack of control of modifiable risk factors in the general population is persistent across Europe.[60,61] An additional complicating factor is the growing number of complex multimorbid patients, and the fact that to date little to no attention has been paid to social determinants of health in the prevention of cardiovascular disease.[62]

A paradigm shift in the organisation of healthcare is needed to successfully reduce the burden of cardiovascular disease, and stroke in particular, in women on a population-wide scale. Population health management is a concept that could drive this paradigm shift, since it defines healthcare models along the individuals' continuum of health and well-being, integrating services in health care, prevention, social care and welfare.[63] Tailoring appropriate preventive policies to subgroups in whom care gaps have been identified and the highest potential gains are expected – also known as panel management – enables precision prevention of cardiovascular disease and stroke in particular.[64] The leading analytic framework for population health management identifies population-based risk stratification as one of the essential steps.[63] The EHR-derived risk prediction models presented in **chapters 15** can support population health management in primary care through the automated identification patient subgroups at high risk of cardiovascular events or stroke in particular. An important limitation of the clinical usefulness of EHR-based prediction models is the relatively low quality of the data from which they are derived.[1] Improving EHR data quality may enable us to base more impactful clinical decisions on the predictions of EHR-derived models, such as prescription of preventive medication. A structural improvement of routine data capture is therefore necessary. This, however, will only occur when the right incentives for healthcare organizations and -professionals are in place, which requires reorganization of the healthcare system. In the last decade, the United States National Academy of Medicine proposed the 'learning healthcare system' as the solution. Its principle is that routinely captured health data can be used for feedback for real time improvement of care processes, through personalization of treatment and diagnosis.[65] This requires the alignment of science, informatics and healthcare

organisations (Figure 5). Large scale examples of learning healthcare systems in the United States such as the Veterans Health Administration Quality Enhancement Research Initiative have shown an acceleration of clinical impact from research in a learning healthcare system.[66] A recent example from the Utrecht Cardiovascular Cohort in the Netherlands shows that an investment in the integration of primary care, hospital care services and data registration has led to improved cardiovascular risk factor registration in about one-third of participants.[67] A learning healthcare system could substantially reduce the 'valley of death' between development and implementation of clinical prediction models. Because of the real time availability of EHR data, prediction model performance and predictor- and outcome distributions could be continuously monitored to detect potential calibration drift.[46]

Figure 5. Schematic overview of a learning healthcare system



The learning healthcare system adapted from Lessard et al.[68], combined with the integration levels of the rainbow model (macro, meso, and micro) for population health management.[69] The blue arrows indicate the flow of data and resources that are necessary for the creation and implementation of a CDSS product for precision prevention of stroke in women.

Not only prediction but also causal inference research may benefit from a learning healthcare system. By increasing the quality of the routinely registered outcome and exposure data in the EHR until it approaches that of traditional cohort research, we could assess outcome-exposure relationships while limiting bias and confounding.[70] It may become the most realistic way to investigate the association between (interactions among) women-specific risk factors and stroke in young women, given the enormous resources (including time) that are needed for traditional prospective cohort research. It currently takes 17 years on average before new knowledge generated by randomised controlled trials is implemented in clinical practice.[71] By embedding pragmatic randomised controlled trials in a learning healthcare system we could potentially validate the effectiveness of preventive policies tailored to subgroups of women at high risk of stroke, to further realise precision prevention while curtailing the costs and total implementation time.

Finally, I present ten recommendations on the system-, organisation- and individual integration level according to the learning population health management concept. These steps may help realise precision prevention of stroke in women in the coming decade. (Table)

Table. Ten recommendations for moving towards a learning healthcare system for precision prevention of stroke in women

Integration level	Recommendation
Macro (system) level E.g. ministry of health, financial bodies, healthcare insurers	1. Organising the right incentives in the healthcare system that reward positive health outcomes instead of financing care based on price per volume.
	2. Implementing healthcare system-wide quality frameworks and -standards ensuring the valid development and implementation of prediction models.
	3. Facilitation of a national data infrastructure which facilitates pooling or joint analysis of data sources.
Meso (organisation) level E.g. healthcare, public-private partnerships, academia	4. Providing education and enabling behaviour change of healthcare professionals.
	5. Fostering productive Public-private partnership for building CDSS products that enable population-based use of clinical prediction models.
	6. Enrichment of routine care data of and optimal information extraction from EHR data.
	7. Increased utilisation of routine care data as extension of evidence based medicine.
Micro (professional) level E.g. [data]scientist, healthcare professional	8. Implementing population health management and empanelment in organisations in healthcare and the social domain.
	9. Co-creation with and behavioural change in the patient and healthcare professional as end-users.
Micro (patient) level	10. Patient empowerment and -participation in data collection.

Ad 1. Value-based healthcare – a healthcare delivery model in which health care providers receive payment based on patient health outcomes – is an essential requirement for population health management, because steering for population health outcomes rather than production can reward the integration of services across healthcare, prevention, and social care services, and the organisation of care across the echelons of primary, secondary and tertiary care.[72] To move towards a learning population health management system, it is also important to incentivise the capture and management of high quality routine care data.

Ad 2. As of 2021, in Europe clinical prediction models for diagnosis or prognosis are classified as medical devices according to the Medical Devices Regulation.

However, this legal framework does not require the methodological rigour that is needed for peer-reviewed validation studies. Therefore, it is important to introduce additional complementary methodological frameworks, which has led to a Dutch field standard for the development and implementation of clinical AI prediction models.[73] Next steps include embedding this field norm in the healthcare system to prevent harmful results of invalid prediction model results.

Ad 3. Pooling or joint analysis of different data sources allows for better phenotyping of patient subgroups, which could lead to precision prevention. For example, socioeconomic and ethnicity data are important predictors for cardiovascular risk and can explain sex differences to some extent.[74] However, the highest possible standards for privacy should be maintained. An alternative to pooling data sources is federated learning, which is based on algorithms that 'visit' data sources and are updated by information without the need of actually pooling of data.[75]

Ad 4. The successful implementation of a CDSS requires education of all end users in the correct use of such systems, and in on underlying epidemiological assumptions and the probabilistic nature of the CDSS, and the relevant legal framework.[73] A learning population health system would also require all end users to know the importance of a high quality routine data collection, and accordingly education of end users is needed on data standardisation and other quality standards.

Ad 5. In public-private partnerships, academic knowledge and access to clinical care can be combined with resources from the private sector, increasing the likelihood of successful implementation and upscaling of CDSS in clinical care.[76] A key driver of a successful public-private partnership is a good business model.[77] Both academic and healthcare organisations, and digital health manufacturing companies should therefore invest in joint business cases for CDSS implementations that enable precision prevention in clinical practice.

Ad 6. A major challenge in using EHR data to derive prediction models for women under 50 years is the extent of missing values in predictors such as systolic blood pressure and smoking. Several studies show that coded routine data can be further enriched by extracting clinical concepts from free text entered into the EHR for example during consultations[78], or by extracting ICD-10 diagnosis codes from discharge letters.[79] For example, smoking can be identified from text in the EHR with 88% sensitivity and 92% specificity.[80]

Ad 7. More should be invested in methodological guidelines about how EHR-derived cohorts can complement traditional (registry-based) research cohorts in answering aetiological questions or developing prediction models.[81] In particular, guidelines should further specify how to deal with the limited quality of routine care

data in aetiological or prediction research. The RECORD statement already provides a first step for this.[82] Further development of methodologies for quantitative bias assessment could help us interpret EHR-based model predictions in the context of the many data quality issues.[83]

Ad 8. Healthcare organisations should organise care paths along the continuum of health and well-being of the individual, which implies the optimisation of care across the echelons of primary, secondary and tertiary care. Panel management may also allow for more efficient task differentiation in primary care, which could increase effectiveness of care and help manage workload.[84]

Ad 9. The primary end users of a CDSS for the risk of cardiovascular events or stroke specifically will be healthcare professionals and patients, and a real impact on health outcomes will only be achieved through behaviour change among these end users. Their involvement from the beginning of the developmental process is a critical factor in the success of the implementation of the CDSS.[73] Around 70% of cardiovascular risk is caused by modifiable risk factors.[59] Therefore, improving healthy behaviour in patients who are at high risk of cardiovascular disease is paramount.

Ad 10. Empowerment of patients – helping them to discover the inherent capacity to be responsible for one's own health – may be a critical modifying factor for the effect of the implementation of a CDSS in clinical practice on lifestyle change.[85] It is, however, important to take cultural and (health) literacy barriers to empowerment into account.[86] Patient empowerment may also result in patients taking a more active role in data collection, either via digital questionnaires with patient-reported outcomes, or via wearable data streams such as digital blood pressure monitors.[87]

References

1. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J.* 2014;35:1925-1931
2. Cook NR, Buring JE, Ridker PM. The effect of including C-reactive protein in cardiovascular risk prediction models for women. *Ann Intern Med.* 2006;145:21-29
3. Singh GM, Danaei G, Farzadfar F, Stevens GA, Woodward M, Wormser D, Kaptoge S, Whitlock G, Qiao Q, Lewington S, et al. The age-specific quantitative effects of metabolic risk factors on cardiovascular diseases and diabetes: a pooled analysis. *PLoS One.* 2013;8:e65174
4. Kurth T, Schurks M, Logroscino G, Gaziano JM, Buring JE. Migraine, vascular risk, and cardiovascular events in women: prospective cohort study. *BMJ.* 2008;337:a636
5. Wu P, Haththotuwa R, Kwok CS, Babu A, Kotronias RA, Rushton C, Zaman A, Fryer AA, Kadam U, Chew-Graham CA, et al. Preeclampsia and Future Cardiovascular Health: A Systematic Review and Meta-Analysis. *Circ Cardiovasc Qual Outcomes.* 2017;10
6. Bellamy L, Casas JP, Hingorani AD, Williams DJ. Pre-eclampsia and risk of cardiovascular disease and cancer in later life: systematic review and meta-analysis. *BMJ.* 2007;335:974
7. Dreier JP, Kremer C, Lammers G, Lohmann F, Hansen HC, Valdueza JM. Migraine and delayed ischaemic neurological deficit after subarachnoid haemorrhage in women: a case-control study. *Eur J Neurol.* 2007;14:1363-1368
8. Jewell NP. Small-sample bias of point estimators of the odds ratio from matched sets. *Biometrics.* 1984;40:421-435
9. Headache Classification Committee of the International Headache S. The International Classification of Headache Disorders, 3rd edition (beta version). *Cephalalgia.* 2013;33:629-808
10. van der Willik D, Pelzer N, Algra A, Terwindt GM, Wermer MJ. Assessment of Migraine History in Patients with a Transient Ischemic Attack or Stroke; Validation of a Migraine Screener for Stroke. *Eur Neurol.* 2016;77:16-22
11. Mahmoud AN, Mentias A, Elgendy AY, Qazi A, Barakat AF, Saad M, Mohsen A, Abuzaid A, Mansoor H, Mojadidi MK, et al. Migraine and the risk of cardiovascular and cerebrovascular events: a meta-analysis of 16 cohort studies including 1 152 407 subjects. *BMJ Open.* 2018;8:e020498
12. Adelborg K, Szepligeti SK, Holland-Bill L, Ehrenstein V, Horvath-Puho E, Henderson VW, Sorensen HT. Migraine and risk of cardiovascular diseases: Danish population based matched cohort study. *BMJ.* 2018;360:k96
13. Launer LJ, Terwindt GM, Ferrari MD. The prevalence and characteristics of migraine in a population-based cohort: the GEM study. *Neurology.* 1999;53:537-542
14. Stewart WF, Linet MS, Celentano DD, Van Natta M, Ziegler D. Age- and sex-specific incidence rates of migraine with and without visual aura. *Am J Epidemiol.* 1991;134:1111-1120.

15. Lipton RB, Bigal ME, Diamond M, Freitag F, Reed ML, Stewart WF, Group AA. Migraine prevalence, disease burden, and the need for preventive therapy. *Neurology*. 2007;68:343-349
16. Lipton RB, Stewart WF, Diamond S, Diamond ML, Reed M. Prevalence and burden of migraine in the United States: data from the American Migraine Study II. *Headache*. 2001;41:646-657
17. Lipton RB, Scher AI, Kolodner K, Liberman J, Steiner TJ, Stewart WF. Migraine in the United States: epidemiology and patterns of health care use. *Neurology*. 2002;58:885-894
18. Goldstein BA, Navar AM, Pencina MJ, Ioannidis JP. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. *J Am Med Inform Assoc*. 2017;24:198-208
19. Kuiper JG, Bakker M, Penning-van Beest FJA, Herings RMC. Existing Data Sources for Clinical Epidemiology: The PHARMO Database Network. *Clin Epidemiol*. 2020;12:415-422
20. Oie LR, Kurth T, Gulati S, Dodick DW. Migraine and risk of stroke. *J Neurol Neurosurg Psychiatry*. 2020;91:593-604
21. Meng XL. Statistical paradises and paradoxes in big data (I): Law of large populations bdp, and the 2016 US presidential election. 2018. The Annals of Applied Statistics. 12:685-726
22. Miettinen O. Estimability and estimation in case-referent studies. *Am J Epidemiol*. 1976;103:226-235
23. Kurth T, Winter AC, Eliassen AH, Dushkes R, Mukamal KJ, Rimm EB, Willett WC, Manson JE, Rexrode KM. Migraine and risk of cardiovascular disease in women: prospective cohort study. *BMJ*. 2016;353:i2610
24. Altman N, Krzywinski M. The curse(s) of dimensionality. *Nat Methods*. 2018;15:399-400
25. Hernan MA, Cole SR. Invited Commentary: Causal diagrams and measurement bias. *Am J Epidemiol*. 2009;170:959-962; discussion 963-954
26. Pajouheshnia R, van Smeden M, Peelen LM, Groenwold RHH. How variation in predictor measurement affects the discriminative ability and transportability of a prediction model. *J Clin Epidemiol*. 2019;105:136-141
27. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, Pencina MJ, Kattan MW. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. 2010;21:128-138
28. Damen JA, Hooft L, Schuit E, Debray TP, Collins GS, Tzoulaki I, Lassale CM, Siontis GC, Chiocchia V, Roberts C, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. *BMJ*. 2016;353:i2416
29. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW, Topic Group 'Evaluating diagnostic t, prediction models' of the Si. Calibration: the Achilles heel of predictive analytics. *BMC Med*. 2019;17:230

30. Steyerberg E.W. (2019) Overfitting and Optimism in Prediction Models. In: Clinical Prediction Models. Statistics for Biology and Health. Springer.
31. Concato J, Peduzzi P, Holford TR, Feinstein AR. Importance of events per independent variable in proportional hazards analysis. I. Background, goals, and general strategy. *J Clin Epidemiol.* 1995;48:1495-1501. doi
32. van der Ploeg T, Austin PC, Steyerberg EW. Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. *BMC Med Res Methodol.* 2014;14:137
33. Ambroise C, McLachlan GJ. Selection bias in gene extraction on the basis of microarray gene-expression data. *Proc Natl Acad Sci U S A.* 2002;99:6562-6566
34. Belkin M, Hsu D, Ma S, Mandal S. Reconciling modern machine-learning practice and the classical bias-variance trade-off. *Proc Natl Acad Sci U S A.* 2019;116:15849-15854
35. Christodoulou E MJ, Collins GS, Steyerberg EW, Verbakel JY, Van Calster, B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. JCE. 2019;110:12-22
36. Brugnara G, Neuberger U, Mahmutoglu MA, Foltyn M, Herweh C, Nagel S, Schonenberger S, Heiland S, Ulfert C, Ringleb PA, et al. Multimodal Predictive Modeling of Endovascular Treatment Outcome for Acute Ischemic Stroke Using Machine-Learning. *Stroke.* 2020;51:3541-3551
37. van Diepen M, Ramspek CL, Jager KJ, Zoccali C, Dekker FW. Prediction versus aetiology: common pitfalls and how to avoid them. *Nephrol Dial Transplant.* 2017;32:ii1-ii5
38. Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *BMJ.* 2009;338:b375
39. Smith GD, Phillips AN. Confounding in epidemiological studies: why "independent" effects may not be all they seem. *BMJ.* 1992;305:757-759
40. Davis SE, Lasko TA, Chen G, Siew ED, Matheny ME. Calibration drift in regression and machine learning models for acute kidney injury. *J Am Med Inform Assoc.* 2017;24:1052-1061
41. Davis, S. E., Lasko, T. A., Chen, G., & Matheny, M. E. (2018). Calibration Drift Among Regression and Machine Learning Models for Hospital Mortality. AMIA ... Annual Symposium proceedings. AMIA Symposium, 2017, 625–634
42. Wells BJ, Chagin KM, Nowacki AS, Kattan MW. Strategies for handling missing data in electronic health record derived data. *EGEMS (Wash DC).* 2013;1:1035
43. Agniel D, Kohane IS, Weber GM. Biases in electronic health record data due to processes within the healthcare system: retrospective observational study. *BMJ.* 2018;361:k1479
44. van Smeden M, Groenwold RHH, Moons KG. A cautionary note on the use of the missing indicator method for handling missing data in prediction research. *J Clin Epidemiol.* 2020;125:188-190
45. Praktische handleiding bij de NHG-Standaard CVRM (2019). Nederlands Huisartsen Genootschap CP, Kwaliteit en Innovatie. Versie 2.1 Juli 2020

46. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med.* 2019;17:195
47. Hu SC, Lan CE. Psoriasis and Cardiovascular Comorbidities: Focusing on Severe Vascular Events, Cardiovascular Risk Factors and Implications for Treatment. *Int J Mol Sci.* 2017;18
48. Terman SW, Aubert CE, Hill CE, Skvarce J, Burke JF, Mintzer S. Cardiovascular disease risk, awareness, and treatment in people with epilepsy. *Epilepsy Behav.* 2021;117:107878
49. Oertelt-Prigione S, Seeland U, Kendel F, Rucke M, Floel A, Gaissmaier W, Heim C, Schnabel R, Stangl V, Regitz-Zagrosek V. Cardiovascular risk factor distribution and subjective risk estimation in urban women--the BEFRI study: a randomized cross-sectional study. *BMC Med.* 2015;13:52
50. Cooney MT, Vartiainen E, Laatikainen T, De Bacquer D, McGorrian C, Dudina A, Graham I, Score, investigators F. Cardiovascular risk age: concepts and practicalities. *Heart.* 2012;98:941-946
51. Graham IM, Di Angelantonio E, Visseren F, De Bacquer D, Ference BA, Timmis A, Halle M, Vardas P, Huculeci R, Cooney MT, et al. Systematic Coronary Risk Evaluation (SCORE): JACC Focus Seminar 4/8. *J Am Coll Cardiol.* 2021;77:3046-3057
52. Kappen TH, van Klei WA, van Wolfswinkel L, Kalkman CJ, Vergouwe Y, Moons KGM. Evaluating the impact of prediction models: lessons learned, challenges, and recommendations. *Diagn Progn Res.* 2018;2:11
53. Score Working Group. SCORE2 risk prediction algorithms: new models to estimate 10-year risk of cardiovascular disease in Europe. *Eur Heart J.* 2021;42:2439-2454
54. Gillman MW, Hammond RA. Precision Treatment and Precision Prevention: Integrating "Below and Above the Skin". *JAMA Pediatr.* 2016;170:9-10
55. Jaspers NEM, Ridker PM, Dorresteijn JAN, Visseren FLJ. The prediction of therapy-benefit for individual cardiovascular disease prevention: rationale, implications, and implementation. *Curr Opin Lipidol.* 2018;29:436-444
56. Powell BJ, McMillen JC, Proctor EK, Carpenter CR, Griffey RT, Bunger AC, Glass JE, York JL. A compilation of strategies for implementing clinical innovations in health and mental health. *Med Care Res Rev.* 2012;69:123-157
57. Camacho J, Zanoletti-Mannello M, Landis-Lewis Z, Kane-Gill SL, Boyce RD. A Conceptual Framework to Study the Implementation of Clinical Decision Support Systems (BEAR): Literature Review and Concept Mapping. *J Med Internet Res.* 2020;22:e18388
58. Vogel B, Acevedo M, Appelman Y, Bairey Merz CN, Chieffo A, Figtree GA, Guerrero M, Kunadian V, Lam CSP, Maas A, et al. The Lancet women and cardiovascular disease Commission: reducing the global burden by 2030. *Lancet.* 2021
59. Yusuf S, Joseph P, Rangarajan S, Islam S, Mente A, Hystad P, Brauer M, Kutty VR, Gupta R, Wielgosz A, et al. Modifiable risk factors, cardiovascular disease, and mortality

in 155 722 individuals from 21 high-income, middle-income, and low-income countries (PURE): a prospective cohort study. *Lancet*. 2020;395:795-808

60. Naderi SH BJ, Wald DS. Adherence to drugs that prevent cardiovascular disease: meta-analysis on 376,162 patients. Am J Med. 2012 Sep;125(9):882-7.e1. doi: 10.1016/j.amjmed.2011.12.013. Epub 2012 Jun 27. PMID: 22748400.
61. Guallar E, Banegas JR, Blasco-Colmenares E, Jimenez FJ, Dallongeville J, Halcox JP, Borghi C, Masso-Gonzalez EL, Tafalla M, Perk J, et al. Excess risk attributable to traditional cardiovascular risk factors in clinical practice settings across Europe - The EURIKA Study. *BMC Public Health*. 2011;11:704
62. C. N. Integrating a Population Health Approach into Healthcare Service Delivery and Decision Making. Healthcare Management Forum. 2012;25(3):155-159
63. Struijs JN, Drewes HW, Heijink R, Baan CA. How to evaluate population management? Transforming the Care Continuum Alliance population health guide toward a broadly applicable analytical framework. *Health Policy*. 2015;119:522-529
64. Khoury MJ, Galea S. Will Precision Medicine Improve Population Health? *JAMA*. 2016;316:1357-1358
65. Friedman C, Rubin J, Brown J, Buntin M, Corn M, Etheredge L, Gunter C, Musen M, Platt R, Stead W, et al. Toward a science of learning systems: a research agenda for the high-functioning Learning Health System. *J Am Med Inform Assoc*. 2015;22:43-50
66. Kilbourne AM, Elwy AR, Sales AE, Atkins D. Accelerating Research Impact in a Learning Health Care System: VA's Quality Enhancement Research Initiative in the Choice Act Era. *Med Care*. 2017;55 Suppl 7 Suppl 1:S4-S12
67. Groenhof TKJ, Lely AT, Haitjema S, Nathoe HM, Kortekaas MF, Asselbergs FW, Bots ML, Hollander M, group UCs. Evaluating a cardiovascular disease risk management care continuum within a learning healthcare system: a prospective cohort study. *BJGP Open*. 2020;4
68. Lessard L, Michalowski W, Fung-Kee-Fung M, Jones L, Grudniewicz A. Architectural frameworks: defining the structures for implementing learning health systems. *Implement Sci*. 2017;12:78
69. Valentijn PP, Schepman SM, Opheij W, Bruijnzeels MA. Understanding integrated care: a comprehensive conceptual framework based on the integrative functions of primary care. *Int J Integr Care*. 2013;13:e010
70. Hernán MA, Robins JM. Causal Inference: What If. 2020. Boca Raton: Chapman & Hall/CRC
71. (US) Institutes of Medicine. In: *Crossing the Quality Chasm: A New Health System for the 21st Century*. 2001.
72. Porter ME. What Is Value in Health Care? N Engl J Med 2010; 363:2477-2481. doi:
73. de Hond AAH, Leeuwenberg AM, Hooft L, Kant IMJ, Nijman SWJ, van Os HJA, Aardoom JJ, Debray TPA, Schuit E, van Smeden M, et al. Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review. *NPJ Digit Med*. 2022;5:2

74. Kist JM, Smit GWG, Mairuhu ATA, Struijs JN, Vos RC, van Peet PG, Vos HMM, Beishuizen ED, Sijpkens YWJ, Groenwold RHH, et al. Large health disparities in cardiovascular death in men and women, by ethnicity and socioeconomic status in an urban based population cohort. *EClinicalMedicine*. 2021;40:101120
75. van Egmond MB, Spini G, van der Galien O, A IJ, Veugen T, Kraaij W, Sangers A, Rooijakkers T, Langenkamp P, Kamphorst B, et al. Privacy-preserving dataset combination and Lasso regression for healthcare predictions. *BMC Med Inform Decis Mak*. 2021;21:266
76. Torchia M CA, Morner M. Public–Private Partnerships in the Health Care Sector: A systematic review of the literature. 2015. Public Management Review, 17:2, 236-261,.
77. van Limburg M, van Gemert-Pijnen JE, Nijland N, Ossebaard HC, Hendrix RM, Seydel ER. Why business modeling is crucial in the development of eHealth technologies. *J Med Internet Res*. 2011;13:e124
78. Wang Z, Shah AD, Tate AR, Denaxas S, Shawe-Taylor J, Hemingway H. Extracting diagnoses and investigation results from unstructured text in electronic health records by semi-supervised machine learning. *PLoS One*. 2012;7:e30412
79. Bagheri AS, A; Van der Heijden, PGM; Asselbergs, FW; Oberski, DL; (2020) Automatic ICD-10 classification of diseases from Dutch discharge letters. In: Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 3: C2C. (pp. pp. 281-289)
80. Groenhof TKJ, Koers LR, Blasse E, de Groot M, Grobbee DE, Bots ML, Asselbergs FW, Lely AT, Haitjema S, Upod, et al. Data mining information from electronic health records produced high yield and accuracy for current smoking status. *J Clin Epidemiol*. 2020;118:100-106
81. Hemkens LG, Contopoulos-Ioannidis DG, Ioannidis JPA. Routinely collected data and comparative effectiveness evidence: promises and limitations. *CMAJ*. 2016;188:E158-E164
82. Nicholls SG, Quach P, von Elm E, Guttmann A, Moher D, Petersen I, Sorensen HT, Smeeth L, Langan SM, Benchimol EI. The REporting of Studies Conducted Using Observational Routinely-Collected Health Data (RECORD) Statement: Methods for Arriving at Consensus and Developing Reporting Guidelines. *PLoS One*. 2015;10:e0125620
83. Lash TL, Fox MP, MacLehose RF, Maldonado G, McCandless LC, Greenland S. Good practices for quantitative bias analysis. *Int J Epidemiol*. 2014;43:1969-1985
84. Bodenheimer T, Sinsky C. From triple to quadruple aim: care of the patient requires care of the provider. *Ann Fam Med*. 2014;12:573-576
85. Roumie CL, Elasy TA, Greevy R, Griffin MR, Liu X, Stone WJ, Wallston KA, Dittus RS, Alvarez V, Cobb J, et al. Improving blood pressure control through provider education, provider alerts, and patient education: a cluster randomized trial. *Ann Intern Med*. 2006;145:165-175

86. Agner J, Braun KL. Patient empowerment: A critique of individualism and systematic review of patient perspectives. *Patient Educ Couns.* 2018;101:2054-2064

87. Dodge HH ZJ, Mattek NC, Austin D, Kornfeld J, Kaye JA (2015) Use of High-Frequency In-Home Monitoring Data May Reduce Sample Sizes Needed in Clinical Trials. PLoS One 10(9): e0138095