TANDVÅRDS- OCH
LÄKEMEDELSFÖRMÅNSVERKET

Follow-up of drug utilisation and treatment effects in clinical practice

Dental and Pharmaceutical Benefits Agency, December 2018
Authors: Sofie Gustafsson, Pontus Johansson, Johan Pontén, Anna Strömgren and Anders Viberg
Reference number: 03381/2018
Corrected version: 25/01/2019

Postal address: Box 22520, SE-104 22 Stockholm
Visiting address: Fleminggatan 18, Stockholm
Telephone: +46 8 568 420 50
www.tlv.se

Preface

Through the Government's appropriation directions for 2017, the Dental and Pharmaceutical Benefits Agency (TLV) was tasked with two assignments which were to be reported by no later than 31 December 2018. One assignment was to conduct two pilot studies on treatment effect in clinical practice. The other assignment was to work in partnership with the Swedish Association of Local Authorities and Regions (SALAR) to conduct a pilot study with the aim of developing a quality registry for monitoring cancer-related drug utilisation at a national level. Within the scope of these assignments, TLV conducted three additional pilot studies focused on collecting and analysing data for drug utilisation follow-up. This report presents the results from the pilot studies for both Government assignments, as well as the lessons learned through the pilot study work and TLV's other work to develop the utilisation of data from clinical practice. We also describe the need for continued work. A more detailed description of the pilot studies is found in appendices to the report.

It would have not been possible for TLV to carry out this work without assistance. We would therefore like to thank the National Board of Health and Welfare, as well as the county councils, companies, registry holders and academic institutions whose expertise, data and work input made completion of the pilot studies possible.

Sofia Wallström
Director-General

Table of contents

1. Introduction

1.1 Assignments and implementation

Through the Government's appropriation directions for 2017, the Dental and Pharmaceutical Benefits Agency (TLV) was tasked with two assignments related to data from clinical practice. In one, the Government mandated TLV to work in partnership with the Swedish Association of Local Authorities and Regions (SALAR) to conduct a pilot study with the aim of developing a quality registry for monitoring cancer-related drug utilisation at a national level. In the other, TLV was tasked with conducting two pilot studies on treatment effect in clinical practice. Both assignments must be reported no later than 31 December 2018.

During implementation of the two Government assignments, TLV determined that information will come to light during the course of the work that will have positive effects in the other pilot studies. In other words, there is a high degree of positive cross-fertilisation between the pilot studies of the two assignments as well as from other pilot studies that TLV initiated within the context of the work. This relates in particular to the availability of data, which has been a complex issue to manage throughout the mandate period. There are long lead times for getting access to data. Improved structures are needed for how data should best be ordered, produced and provided, and the organisation of the process needs to be more transparent. With this in mind, TLV finds it most expedient to present the two Government assignments in the same report, with links to all of the pilot studies conducted within the framework of the work, in order to create the greatest amount of knowledge and largest number of solutions related to follow-up and treatment effects in clinical practice.

The first assignment relates to follow-up of cancer drug utilisation in clinical practice. In partnership with SALAR and the Regional Cancer Centre (RCC), TVL conducted a two-part pilot study, a national follow-up and a regional follow-up. Working together, common issues were defined. Data extraction from the registry was done in collaboration. The approaches tested in the pilot study (Pilot 6) were to use data from two different cancer registries, namely the National Quality Registry for new drugs and RCC Stockholm-Gotland's Quality Registry for new cancer care drugs (Stockholm Registry). The two drugs that TLV chose to study in clinical practice are Opdivo and Keytruda, both for the indication malignant melanoma. The national registry was established relatively recently and currently contains only a small number of drugs. The regional one is a cancer registry for patients treated in Stockholm and Gotland. It is an older quality registry that contains a greater number of variables, but only covers patients from Stockholm and Gotland.

The second assignment relates to using a method to measure treatment effect in clinical practice. Here, TLV commissioned two pilot studies that apply two different methods.

One pilot study (Pilot 2) builds on the Government assignment reported by the Swedish Agency for Health Technology Assessment and Assessment of Social Services (SBU) in 2016[1]. It aims to test a method based on individual data from a randomised clinical trial being used as a reference group when evaluating a drug's efficacy in clinical practice. To test the method, two pharmaceutical companies provided access to data from the clinical trials of the diabetes drugs Forxiga and Victoza. Data from clinical practice was obtained from the national diabetes registry (NDR) and was linked to data from the National Board of Health and Welfare's national health registry. In the second pilot study (Pilot 3), data from Stockholm County Council's (SLL) registry was used to apply a further method for evaluating treatment effect in clinical practice. The method aims to show how data from clinical practice can be used to create a reference group for evaluating the efficacy of a drug when it is used in clinical practice. This method, which does not require data from the clinical trial, was applied to the drug Entresto that is used for treating heart failure. The method is based on finding patients receiving Entresto and comparable patients receiving another treatment. In this context, comparable patients refers to patients with the same degree of disease, age, gender, socioeconomic conditions, etc., which means that the patient could just as well have received Entresto as the comparison group's treatment.

In order to best answer the questions in the Government assignments and to look forward, TLV conducted not only the pilot studies linked to the two assignments but also three additional pilot studies (Pilot 1, Pilot 4 and Pilot 5), with a focus on collecting and analysing data for drug utilisation follow-up.

In Pilot 1, TLV evaluates use of the drug class PCSK9 inhibitors, which are used to treat high lipid levels. Pilot 4 tests whether it is technically possible to connect data from the prescription registry at the Swedish eHealth Agency with parts of the registry for patient administrative data from the county councils, and to visualise data on the SVEUS analysis platform. The drug Tecfidera, which is used for multiple sclerosis (MS), was evaluated in the project with data from Region Skåne and Region Västra Götaland. Pilot 5 involved identifying possibilities for extracting data from the county councils in a more automated manner via the national service platform. This identification focused on general possibilities and is therefore not linked to any specific drug.

1.2 Development of value-based pricing

In recent years, the Government's appropriation directions to TLV have specified a goal for the Agency to develop value-based pricing in order to ensure that drugs are cost-effective throughout their life cycle, and to promote innovation by promoting the use of new, innovative and cost-effective drugs. The development of value-based pricing takes place, inter alia, through more dynamic pricing, which increases the

[1] SBU, Report no. 256 (2016) https://www.sbu.se/256

requirement for TLV to follow up and evaluate how drugs are used in clinical practice.

Follow-up becomes even more important in light of the increase in the proportion of reimbursement decisions with restrictions, i.e. where the drug has been judged to have a reasonable cost only with use by a limited patient group. Tripartite deliberations[2] that result in side agreements that are added to the case also require developed follow-up.

Good follow-up can make it possible for a greater number of drugs in early phases to be judged to fulfil the requirements for granting reimbursement, which means that patients in Sweden have earlier access to the drugs. At the same time, early knowledge acquisition and access to data on the drug's effects in clinical practice are of great value to the life science industry in Sweden.

The trend in the pharmaceutical field is moving towards the introduction of new drugs with great potential efficacy at an ever earlier stage. Earlier introduction gives patients quicker access to drugs, but at the same time means that one must accept a higher degree of uncertainty about the drug's effects in clinical practice. It is therefore crucial that it is the patients with the greatest medical needs who receive the treatment early. Early introduction of new drugs increases the need to be able to ensure the drug's efficacy in clinical practice, and that the right patient group is using the drug, equally across the country. There is considerable risk that the efficacy seen in clinical practice will differ from that assessed in clinical trials.

Many new drugs require the development of value-based pricing and stimulation of the creation of new business and payment models. These may be individualised therapies based on gene therapy or stem cell treatments, where the treatment is given on a single occasion and is expected to have a lifelong effect. The companies' payment models have not been adapted to this, and many companies are expressing a need for larger compensation on a single occasion, which reduces the possibility of follow-up and completely removes the possibility of ending therapy for an individual patient if the effect does not justify the cost. This challenge, however, is not within the scope of this work. But, the development of methods for the follow-up of drugs in clinical practice is important as it will help us handle this complexity in the future.

In this context, TLV sees a need to obtain detailed knowledge of the market situation for different types of drugs in the same therapy area as regards aspects such as treatment tradition, duration of treatment, new initiation rate, changed use, and data on efficacy in clinical practice. Such knowledge is necessary and can be generated from registry data from both the National Board of Health and Welfare's drug and patient registries (where utilisation of drugs can be analysed relatively

2 Tripartite deliberations - A tripartite deliberation is a joint dialog between the regions, companies and TLV regarding a pharmaceutical assessment. These deliberations can result in a national agreement between the company and the regions. TLV is not a part of the agreement.

well) and from various quality registries (where there is often a greater number of data variables, such as cause of therapy change, actual outcome measures and health effects). The questions that need to be addressed based on these data sources are often complex and require detailed analysis of data at the individual level. The purpose of such analyses is e.g. to create support for decisions on possible alternatives in a review within a therapy area or to provide underlying data on existing treatment before a new drug is evaluated.

In view of this, TLV has a great and growing need to be able to monitor drug utilisation and treatment effects in clinical practice, particularly for new, innovative and often costly drugs. TLV needs for follow-up to take place in close connection to introduction of drug treatment, and for follow-up to be continuous. These needs cannot be met by research in the field. However, close collaboration with researchers is valuable for TLV as the research complements the iterative follow-ups needed when a new drug is introduced. At the same time, TLV needs to access information about drug utilisation in clinical practice, often at a detailed level, in order to evaluate the benefit of the drug, which is linked to what can be considered a reasonable cost for society to pay for the drug. Such information can be used as a basis in TLV's price and reimbursement decisions. By extension, the information is also needed to be able to evaluate decisions and thereby achieve more cost-effective use throughout the entire life cycle of the drug. In order to quickly gain access to new knowledge about actual effects of drug treatment, other actors, such as researchers, must have access to relevant study data. Because relevant study data is often not available, it can take a long time for researchers to generate good data. Registry-based studies of big data from the healthcare sector are beginning to emerge as a potentially faster alternative to traditional studies, and thus an important complement. Quality registries play an important role in this work as they in many cases contain extensive health data which is of great value in follow-up.

The challenges surrounding data on efficacy in clinical practice can be broken down into a number of different aspects. The first is how TLV should be able to access data in some form. Another is the development of suitable and feasible methods. What methods should be applied and in what way? And how can the data best be interpreted? TLV finds that, with better coordination of quality registries and other data sources, Sweden has the potential to become one of the leading nations in the world in terms of phase 4/follow-up studies of innovative drugs. Coupled with a high level of basic research, it strengthens Sweden's attractiveness in the life science industry.

Through the two Government assignments presented in this report and the supplementary pilot studies that TLV conducted on its own initiative (also presented in this report), TLV hopes to contribute to deeper understanding of how the methods can be developed. In this context, collaboration with the National Board of Health and Welfare, the quality registry holders and academia is particularly important, as is continued cooperation with the country councils and pharmaceutical companies.

2 Six pilot studies were conducted within the framework of the assignments

As part of the Government assignments, TLV conducted a total of six pilot studies. The pilot studies are briefly presented in this chapter. Pilot 2–6 are described in more detail in separate appendices to this report.

2.1 Pilot 1: Follow-up of PCSK9 inhibitors

Pilot 1 "Follow-up of PCSK9 inhibitors" is intended to follow-up the level of compliance with the reimbursement restriction for PCSK9 inhibitor, and to gain more knowledge about the recurrence of cardiovascular events and/or death for individuals treated with these drugs in clinical practice. The pilot study clearly illustrates the challenges of interpreting and generalising outcomes in clinical practice which may arise as a result of a lack of relevant data.

The PCSK9 inhibitors Repatha and Praluent are approved for the treatment of high cholesterol to reduce the risk of cardiovascular events. At the time of introduction, there was some uncertainty about the long-term effects. This, coupled with a high drug cost, contributed to PCSK9 inhibitors being mainly limited to treatment in secondary prevention after myocardial infarction in patients who have LDL >4 mmol/L despite other cholesterol-lowering treatment. The Swedeheart quality registry tracks patients who have had a myocardial infarction. Among other times, LDL levels are registered 6–10 weeks after infarction and 12–14 months after infarction. By selecting individuals who started PCSK9 inhibitor treatment between the first and second registration in the registry, the LDL level is registered before and after initiation of treatment. Data from Swedeheart was ordered, and a graph of the individuals who could be studied is presented in Figure 1. Since there is only a small number of individuals in this dataset and they are probably not representative of other patients treated with PCSK9 inhibitors in Sweden, the results are not generalisable. However, a majority of the included patients had an LDL level lower than 4 mmol/L before initiation of PCSK9 inhibitor, indicating that compliance with the reimbursement restriction is low among these patients.

Figure 1. LDL before and after initiation of PCSK9 inhibitor. Each individual is connected by a line. The green section is prior to initiation of PCSK9 inhibitor, while the blue section is after initiation of PCSK9 inhibitor. Data from Swedeheart quality registry.

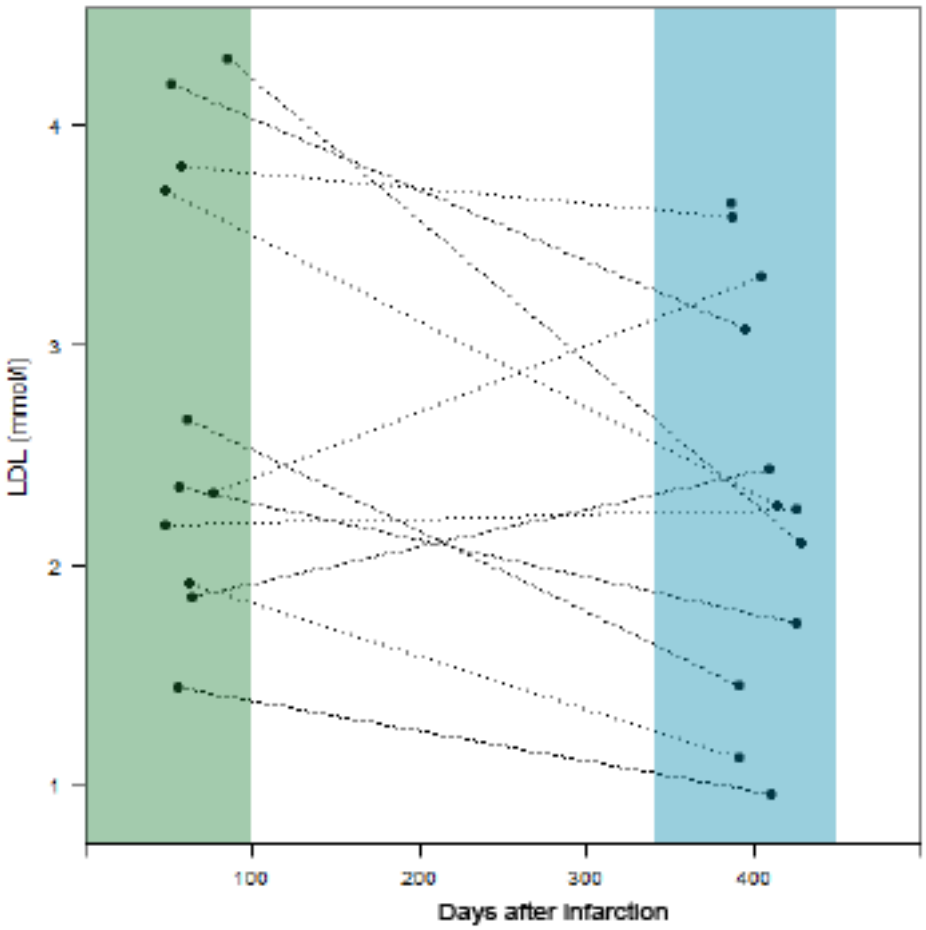
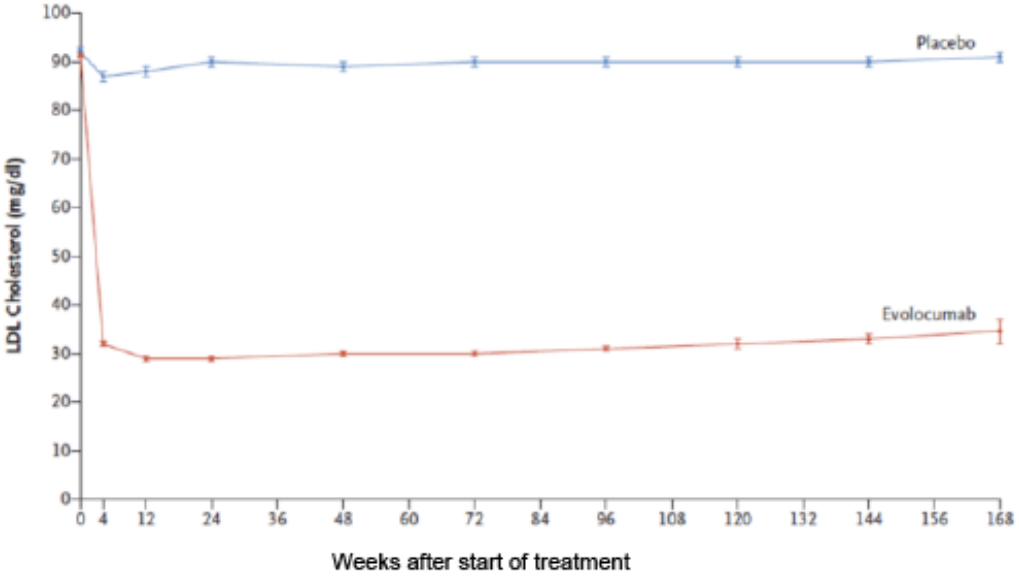


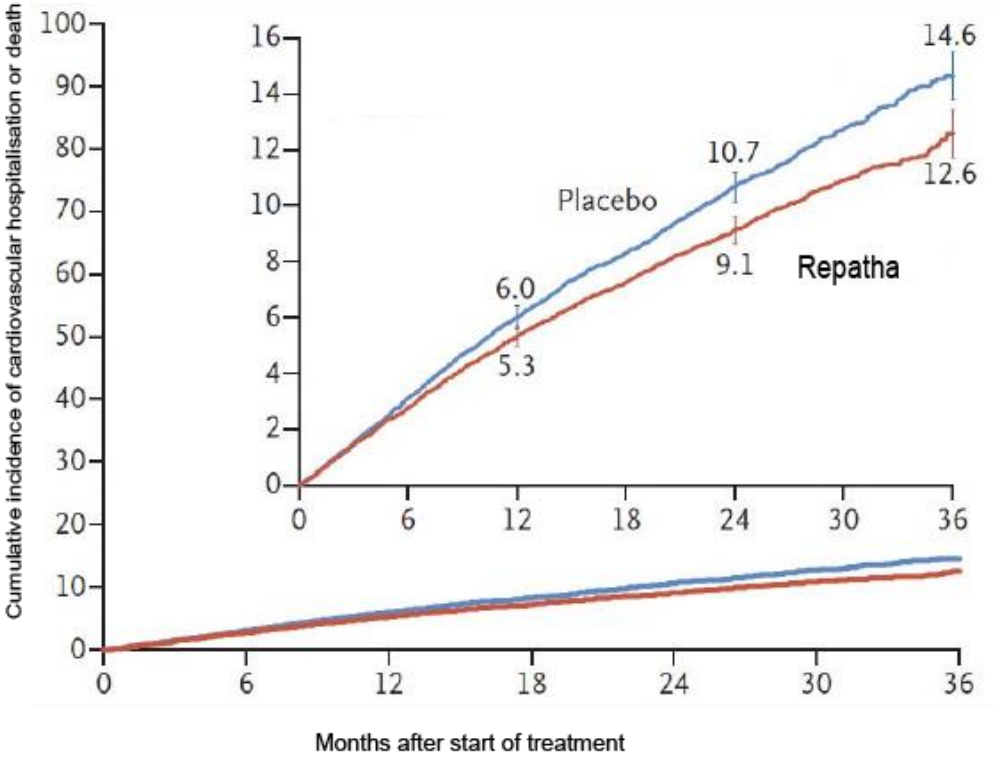
Figure 2 shows how LDL changed as a result of Repatha treatment in the clinical trial that served as the basis for the reimbursement decision. Within as few as four weeks, LDL levels are reduced by almost 70 percent, while data from the quality registry (Figure 1) does not indicate any significant reduction in levels after starting treatment with a PCSK9 inhibitor. Thus, the LDL reduction observed in the clinical trial for Repatha cannot be verified with data from the quality registry.

Figure 2. LDL change versus time after start of treatment. Modified from Sabatine et al. N Engl J Med. 2017 May 4;376(18):1713-1722



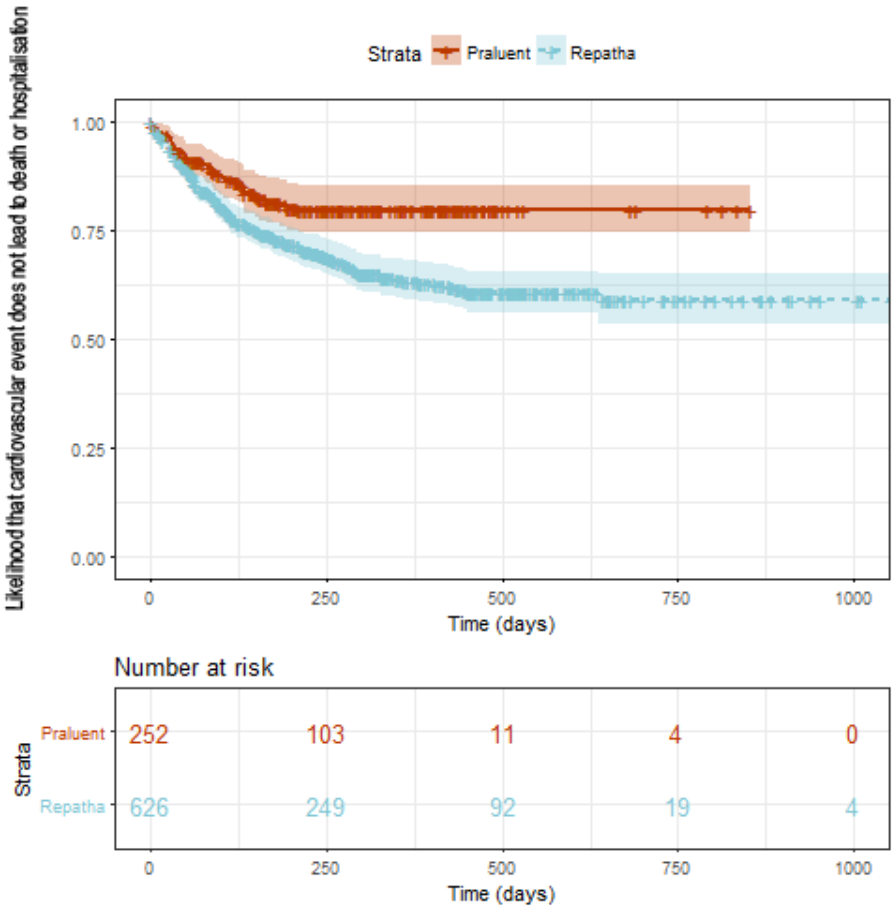
The clinical trial for Repatha showed a significant decrease in hospitalisation or death due to cardiovascular event when treating with Repatha as compared to placebo; see Figure 3.

Figure 3. Incidence of cardiovascular death or hospitalisation. Modified from Sabatine et al. N Engl J Med. 2017 May 4;376(18):1713-1722



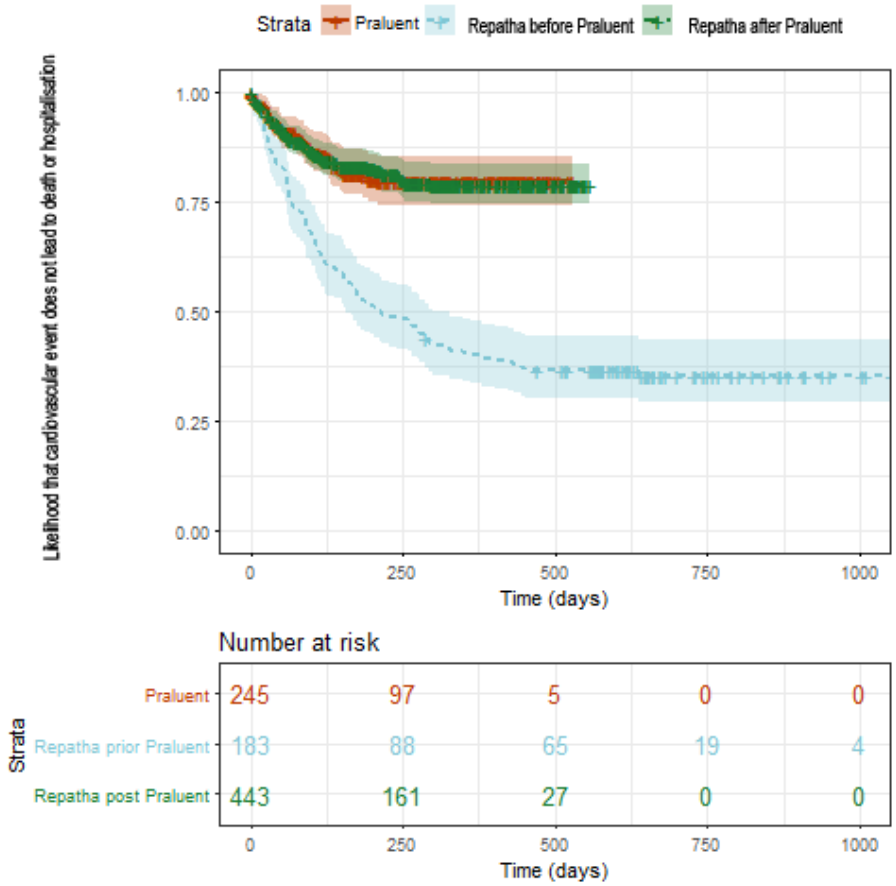
To try to investigate the frequency of cardiovascular events in the group treated with PCSK9 inhibitors in Swedish clinical practice, TLV ordered data from the National Board of Health and Welfare's drug registry and patient registry, with information on all dispensings of the PCSK9 inhibitor Repatha or Praluent. The data contained information on time from first filling of a prescription for PCSK9 inhibitor to any hospitalisation due to cardiovascular event or death. For individuals for whom no additional cardiovascular event was registered, time to the last time point of registered interaction with the healthcare sector or pharmacy was instead used to verify that the patient was still available for evaluation and had not died or moved. This became the time for censoring. Kaplan Meier curves were then created based on this dataset. First, Repatha and Praluent were compared (Figure 4).

Figure 4. Repatha and Praluent time to hospitalisation due to cardiovascular event or death



Based on the Kaplan Meier curves, Repatha appears to have less efficacy than Praluent as the likelihood of hospitalisation or death was higher than that of Praluent. Repatha was granted reimbursement approximately six months earlier than Praluent, which means that Repatha was initially the only PCSK9 inhibitor that was part of the pharmaceutical benefits scheme. To further investigate how this affected the outcome, a variable was included as a control for Repatha being granted reimbursement earlier than Praluent. New Kaplan Meier curves were then created; see Figure 5.

Figure 5. Time to hospitalisation or death for individuals treated with Praluent or Repatha before and after Praluent was included in the pharmaceutical benefits scheme, respectively.



If you compare Repatha and Praluent while taking into account the difference in the time point for reimbursement, we can no longer see any differences in outcome between the two PCSK9 inhibitors. One possible explanation is that patients with the highest risk of cardiovascular disease are treated first. After a while, patients with less risk of cardiovascular disease will also be treated, which over time will increase the proportion of patients at lower risk. The lower risk means that the likelihood of hospitalisation/death is lower even though the drug treatment is the same. When compared to the pivotal phase II study of Repatha, the likelihood of hospitalisation or death is much higher for Repatha in clinical practice compared to both Repatha and placebo treatment in clinical trials. However, this should not be interpreted as meaning that Repatha does not work in clinical practice. It should instead be seen as an indication that the patient population in the clinical trial is less ill than the population treated with Repatha or Praluent in Swedish routine care.

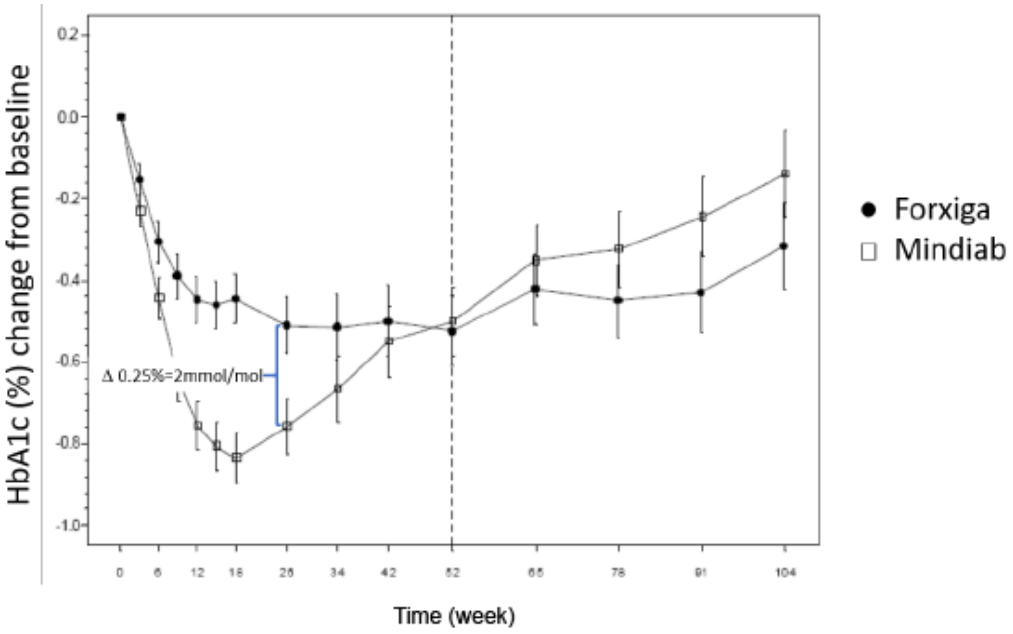
2.2 Pilot 2: Method for evaluating treatment effects in clinical practice with a comparison group from a clinical trial

The purpose of Pilot 2 is to apply the method identified by SBU to measure the efficacy of drugs in clinical practice (real world data, RWD). SBU reported this to the Government in November 2016[3]. The method aims to evaluate treatment effects in clinical practice with a comparison group from a clinical trial. In May 2017, TLV initiated a discussion with four pharmaceutical companies, of which two companies (AstraZeneca and Novo Nordisk) chose to participate in the pilot study. The companies stated that they perceived some ambiguity as to which actors could access individual data from their clinical trials, which contributed to hesitation about participating in the pilot study.

Pilot 2 used individual data from AstraZeneca's clinical trial of Forxiga (dapagliflozin) compared with the sulfonylurea Mindiab (glipizide) and from Novo Nordisk's clinical trial of Victoza which was compared with the sulfonylurea glimepiride. Data from the clinical trials was compared with individual data from the National Diabetes Registry (NDR), which was coordinated at the National Board of Health and Welfare with data from, inter alia, the patient registry and the drug registry. To reduce exposure of the data, the NDR, AstraZeneca and Novo Nordisk uploaded their data themselves to an external data platform. The NDR analysed data via the platform, but did not have the ability to download data.
The primary outcome measure in the clinical trials was change in HbA1c (after 12 months with Forxiga (Figure 6) and after six months with Victoza). The secondary outcome measures were change in body weight and blood pressure. Both clinical trials were non-inferiority studies, which are studies intended to show that the studied drugs are at least as good as the control group. Forxiga and Victoza are therefore not expected to produce a greater reduction in HbA1c. In AstraZeneca's clinical trial, Mindiab lowers HbA1c by approximately 2 mmol/mol more than Forxiga does after six months of treatment (Figure 6), while the reduction is equivalent after 12 months.

[3] SBU, Report no. 256 (2016) https://www.sbu.se/256

Figure 6. HbA1c response in AstraZeneca's study of Forxiga. A difference of 0.25% HbA1c corresponds to a 2 mmol/mol change



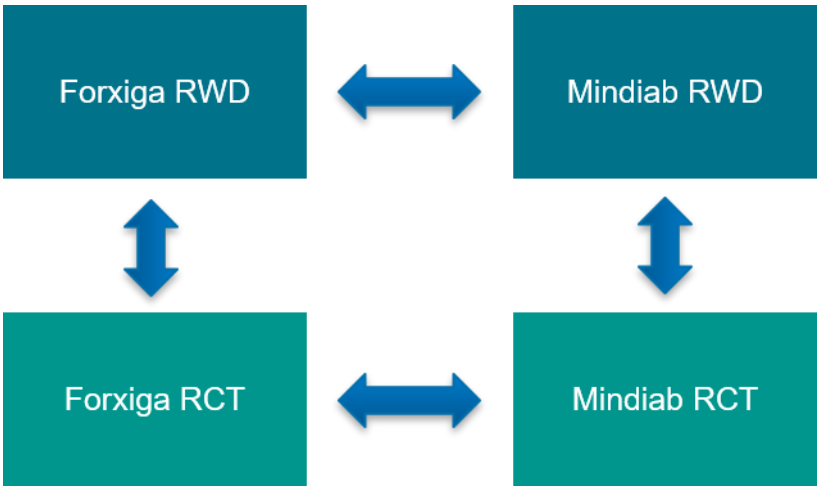
Most of the recorded data in clinical practice occurs at 12-month intervals since it is common for return visits to be annual. For the comparison, the choice was made to compare the response at 6 months of treatment in RCT and in clinical practice. This leads to the exclusion of many individuals in the dataset RWD since the 6-month measurement of HbA1C, body weight and blood pressure are often missing.

2.2.1 Method

The basic premise of the method is to investigate how the efficacy observed in randomised clinical trials (RCT) will scale to efficacy in clinical practice. As you can expect a drug's efficacy to be dependent on the composition of the patient group, you need to be sure that you are comparing the same types of patients in clinical practice, i.e. real world data (RWD), as in the RCT. If they are not the same, you need to compensate for how the demographics can affect the results. This can be addressed by taking into account certain cofactors. In this example, age group, gender, BMI, weight, HbA1c, SBP, DBP, LDL, HDL, triglycerides, creatinine and smoking are available from both randomised trial and clinical practice to enable such weighting of the results. In the randomised trials, the control group and treatment group are the same as assignment of individuals to a treatment is random. In RWD, these two groups must be created by weighting the results for Forxiga or Victoza-treated patients with sulfonylurea-treated patients so that both groups are as equal as possible. In addition, the individuals from RWD must meet the inclusion and exclusion criteria used in RCT. When this is done, you take the results from RCT and predict what efficacy you can expect in clinical practice using the cofactors above. The comparison with the drugs in RCT is already available. Using the AstraZeneca example, you can use RWD to compare Forxiga's efficacy in RWD with the efficacy in RCT. The same comparison can also be made with Mindiab in RCT and RWD. Finally, you can compare the efficacy of Forxiga

compared with Mindiab in RWD. A schematic illustration of this is shown in Figure 7. In both examples, a background treatment of metformin is used with the addition of either the index drug (Forxiga or Victoza) or a sulfonylurea (Mindiab or glimepiride).

Figure 7. The different comparisons that can be made in the evaluation of Forxiga.



2.2.2 Data

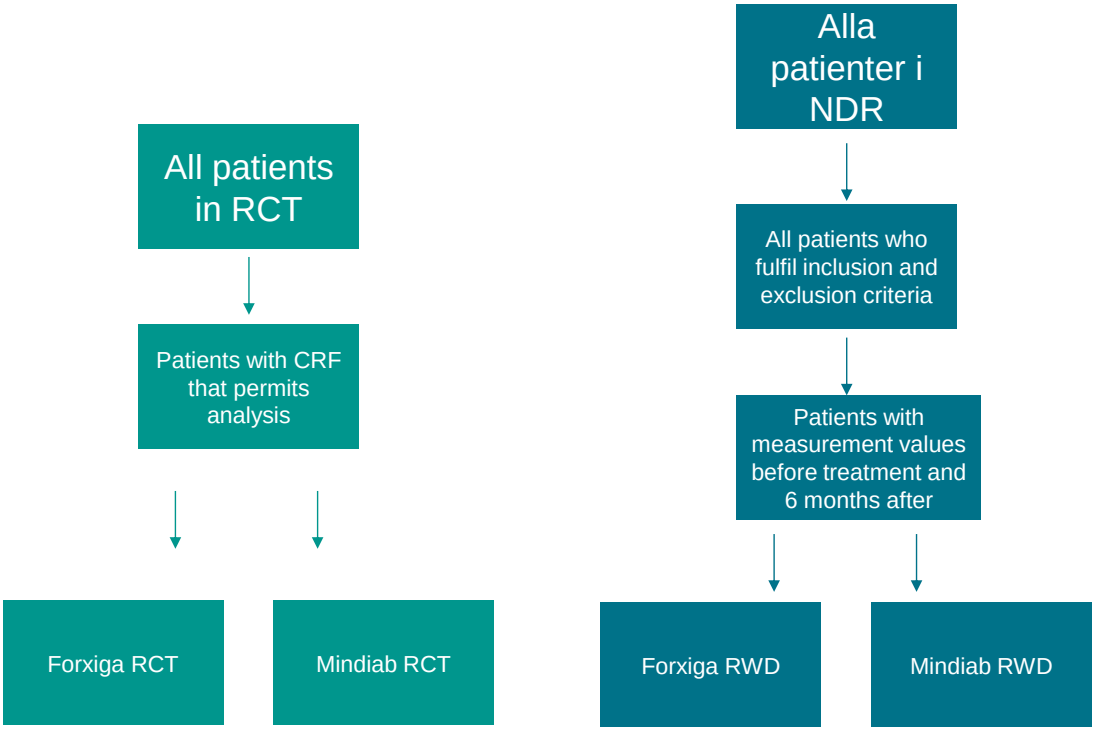
The question was answered through use of data from clinical trials, the National Board of Health and Welfare's drug registry, patient registry and death registry, and Statistics Sweden's (SCB) population register and the National Diabetes Registry. Data was uploaded to an external data platform (without the possibility of download) and analysed by the NDR.

RCT dataset
One RCT dataset consists of AstraZeneca's clinical trial D1690C00004, where Forxiga+metformin is compared with Mindiab+metformin. Origin data from RCT is made up of 814 individuals. Informed consent allowing further analysis existed for 509 individuals, which resulted in RCT data in this study being a subset of the clinical trial.
The other RCT dataset consists of Novo Nordisk trial NN2211-1572, where different doses of Victoza+metformin are compared with glimepiride+metformin. A total of 880 individuals were included.

RWD dataset
The RWD dataset consists of data from the NDR which is linked via the National Board of Health and Welfare to data from the drug registry, patient registry, death registry, and SCB's population register.
In the NDR, a dataset was identified with 46,229 patients who were registered between 1 January 2011 and 31 December 2017 and met the inclusion and exclusion criteria of the Victoza clinical trial and an equivalent dataset with 30,054 patients from 1 June 2013 to 31 December 2017 for Forxiga. Subsequently, individuals were

identified who i) were dispensed Forxiga or Mindiab or dispensed Victoza or glimepiride during the respective time period according to the drug registry, and ii) had HbA1c measurement values prior to initiation of treatment and at 6 months after starting treatment. These latter datasets consisted of 209 patients treated with Forxiga and 1905 with sulfonylurea drugs and 1026 treated with Victoza and 5043 treated with sulfonylurea drugs. A schematic illustration of the flowchart for generating a dataset is shown in Figure 8.

Figure 8. Flow chart of the dataset generated from NDR and RCT patients in AstraZeneca's Forxiga trial.



2.2.3 Results

In the data available from RCT, HbA1c was reduced by approximately 2.2 mmol/mol more after 6 months of treatment with Mindiab compared to 6 months of treatment with Forxiga (Figure 6). When you then use the cofactors described above and apply weighting to accommodate for the difference in composition between the RWD dataset and the RCT dataset, the method can be used to predict the outcome of drug treatments for the population in the RWD dataset. The method predicts that the difference between the treatments in the RWD dataset is slightly larger than that measured in RCT. Mindiab is predicted to lower HbA1c by an average of 2.8 mmol/mol more than Forxiga after 26 weeks of treatment. When you then look at measured HbA1c in the RWD dataset and weight the treatments, you see that the relationship is the opposite, i.e. Forxiga reduces HbA1c by approximately 2.1 mmol/mol *more* than Mindiab after 6 months of treatment.

To understand the impetus behind Forxiga now lowering HbA1c more than Mindiab, the outcome for Forxiga in RCT is now compared with the outcome in clinical practice. When weighted with the patient factors, you can then see that Forxiga lowers HbA1c by 1.1 mmol/mol more in clinical practice compared with RCT. Thus, the outcome is greater in clinical practice than in RCT. For Mindiab, the reverse is true, i.e. HbA1c is lowered by 4.0 mmol/mol more in the clinical trial compared to clinical practice. Forxiga's greater efficacy in clinical practice thus seems attributable to the fact that Mindiab has less effect in clinical practice compared to how it works in RCT rather than Forxiga working better than expected in RWD. Consequently, bridging RCT data and RWD is not straightforward, even if you weight the results based on known patient characteristics. When an equivalent analysis is carried out in relation to body weight, the analysis is more consistent. In RCT, Forxiga reduces weight by approximately 5 kg more than Mindiab does. When weighting RCT to RWD, approximately the same size of effect is predicted. This is ultimately verified by the fact that Forxiga reduces body weight by approximately 4 kg more than Mindiab in clinical practice.

In data from Novo Nordisk, the effect on HbA1c is about the same for Victoza and glimepiride in RCT after 6 months of treatment. When the results are weighted to address the different composition of the populations, it is predicted that the same results should be measured in RWD. In RWD, Victoza lowers HbA1c by approximately 2 mmol/mol more than Mindiab, which is due to a combination of Victoza working slightly better than expected in RWD and glimepiride working slightly worse than expected in RWD. In terms of effect on body weight, the results are the same as for Forxiga, where Victoza reduces weight about as much in RCT as in RWD.

Discussion

Pilot 2 shows that it is possible to use data from clinical trials, clinical practice and other registries to perform a combined analysis to evaluate the efficacy. However, the pilot study shows that using clinical trial data to predict the efficacy observed in clinical practice may vary from case to case. Further method development is needed to identify when it is appropriate to use the method.

Other experiences are that the forms for combined analysis of data between companies and authorities need to be developed. Legal conditions and access to the right analysis tools on the shared analysis platform need further review. Furthermore, data access needs to be improved in relation to outcome data in quality registries and in terms of socioeconomic variables from RCT populations. The lack of data on hypoglycaemia in the quality registry as well as long-term follow-up of more severe outcomes in the respective RCTs is an obstacle to more comprehensive analysis. In other therapy areas with less robust outcomes than HbA1c, additional difficulties may arise with regard to data quality in clinical practice (RWD). However, the positive cooperation climate between companies, registry holders and the authorities shows that there are conditions for developing this methodology and testing it in other areas.

The full sub-report is available here: Pilot 2

Inclusion and exclusion criteria can be found here: Appendix Pilot 2

2.3 Pilot 3: Method for evaluating treatment effects in clinical practice with a comparison group from clinical practice

The purpose of Pilot 3 is to test a method for evaluating treatment effects in clinical practice by using a comparison group that is also taken from clinical practice. In the pilot study, the drug Entresto, which is used to treat heart failure, was used to test the method. The work was conducted in collaboration with Uppsala University.

2.3.1 Method

Creating groups from electronic medical records systems and other healthcare registries where patient groups are equal makes it possible to compare two different treatments with each other. In order to find an equal comparison group, the initial phase of the study placed a great deal of focus on carefully defining study samples, data sources, exposures and comparison exposures. A key task at this stage is developing the causal model, which then serves as the basis for building the model. The model being built must explicitly emulate a hypothetical randomised clinical trial, and the goal is to achieve a situation where the prescribing doctor would have the same probability of initiating the drug in question as initiating the comparison treatment.

In the RCT used when applying for reimbursement for the drug Entresto, an ACE inhibitor was used as a comparator for Entresto. In the introductory phase of the pilot study, when discussing possible comparison groups for Entresto in clinical practice, TLV and the pilot study working group determined that an ACE inhibitor is not a relative comparison group. This is based on the assumption that, during the observation time in question, most doctors probably prescribed Entresto to patients in Sweden who needed intensified treatment despite ongoing treatment with the highest tolerable dose of, as a minimum, the combination of beta blocker and ACE inhibitor/ARB. A more relevant comparison group for Entresto in clinical practice is instead patients with the highest tolerable dose of beta blocker, an ACE inhibitor/ARB, an aldosterone antagonist plus the need for additional add-on therapy in the form of either ivabradine, digoxin or ICD/CRT (implantable defibrillator/pacemaker). In the work with the causal model, a large number of factors were identified that could potentially affect a doctor's choice of treatment or the outcome of treatment. These were used in the statistical model to make the treatment groups as comparable as possible. The primary outcome measure in the comparison was time from start of treatment to either hospitalisation due to heart failure or death (regardless of reason). The index date for comparison between the groups was either the time of Entresto prescription filling or when ivabradine or digoxin was collected or an ICD/CRT inserted.

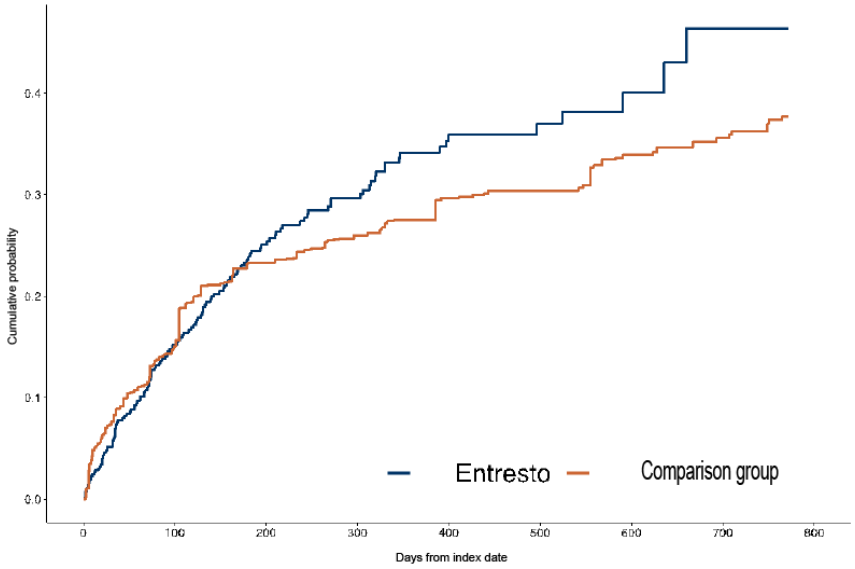
2.3.2 Data

Pilot 3 used data from SLL. Both information from electronic medical records systems and information from regional follow-up databases were used. A request for access to data was addressed to a total of seven different healthcare providers within SLL. An individual application, an individual confidentiality assessment and interaction with decision makers was required for each healthcare provider, and it took seven months to gain access to research data.

2.3.3 Results

In total, the study included 539 patients who began treatment with Entresto and 556 comparison patients, who began treatment with digoxin, ivabradine or CRT/ICD. Good comparability between the groups could be achieved in the form of balance of covariates. The incidence rate in the primary efficacy outcome (first event of outpatient heart failure or death) was similar in both groups; hazard ratio 1.12 (95% confidence interval 0.80–1.58). A comparison between Entresto and the control group is shown in Figure 9.

Figure 9. Kaplan Meier curve for Entresto and the comparison group on the probability of achieving the study's primary outcome; time to first hospital care for heart failure or death.



2.3.4 Discussion

In this pilot study, which compared all patients in SLL who began treatment with Entresto during the study period with similar patients who began other treatments, no definite difference could be seen between the groups in terms of the efficacy or safety or the treatments, according to the study's definitions. The method was able to use data on drug treatment given in outpatient and inpatient care, data on surgical procedures, and data from clinical physiology and clinical chemistry in order to define inclusion criteria, comparison groups, covariates and outcome events. The pilot study had to make a few deviations from the initial statistical

analysis plan as some data was of low quality, which underscored the need for access to relevant data during the initial phase in this model.

The full sub-report is available here: Pilot 3

2.4 Pilot 4: Drug utilisation follow-up on the Sveus analysis platform

Pilot 4 aims to link data from the county councils'/regions' patient administration systems (PAS data) with data from the prescription registry in order to monitor the use of drugs on the Sveus analysis platform.

2.4.1 Method

Pilot 4 was conducted with data from Region Skåne and Region Västra Götaland. The Sveus analysis platform is a platform where data from county councils'/regions' patient administration systems and relevant quality registries is analysed in order to monitor the healthcare provided. Data is kept separate for each county council/healthcare provider, but since analyses are done with the same data structure, comparisons can be made between county councils and these can be aggregated into a combined analysis if necessary. The analysis currently processes data for two-thirds of Sweden's population. The pilot study also aims to illustrate how descriptive statistics on drug utilisation follow-up can be made available in an interactive way via a web interface. Tecfidera for the treatment of MS was chosen as an example for this pilot study.

At the start of the pilot study, five county councils and regions were connected to the Sveus analysis platform (SLL, Region Västra Götaland, Region Skåne, Uppsala and Dalarna). At start-up, all connected regions were asked whether they wanted to participate in the pilot study. At that time, Region Västra Götaland and Region Skåne chose to participate. This was judged by Sveus to be sufficient for a pilot study as it included two major regions, which would not only provide sufficient data base but also enable comparison between two regions. The region-internal decisions to participate in the pilot study were made by the regional steering groups for Sveus.

2.4.2 Data

The purpose of the pilot study was narrow and limited to testing the possibilities of linking the county councils' patient administration systems with prescribed drugs as relates to Tecfidera, and to create an interactive interface to make data accessible without disclosing sensitive data. The experiences from this pilot study are similar to those of the other pilot studies. It takes time to collect data. It is not the actual data extraction and sending of data that is time-consuming. It is the processes for obtaining permits at different levels that take time, both legal processes as regards personal data processing agreements and technological processes e.g. linked to encryption keys. Moreover, these processes differ from one county council to another. It initially takes longer when a new county council joins, while repeating a data delivery for a county council that is already connected should be faster. But, the permits that exist apply to this particular pilot study only. New decisions and

agreements are required in order to carry out new follow-ups or expand this limited follow-up. In this pilot study, data was processed by Ivbar, and the results of the analyses were visualised via the Sveus analysis platform. Ivbar is a research and development company that provides project management of analysis and development work within Sveus.

2.4.3 Results and discussion

In this pilot study, the sample was MS patients treated with the drug Tecfidera. From TLV's perspective, it would have been more relevant for the sample to be all MS patients and how Tecfidera relates to other MS treatments. However, the pilot study focused more on the technical aspects. The results of the pilot study show that it is technically possible to connect PAS data for patients that filled a prescription for the drug Tecifidera, create different key figures, and make these available via a web interface.

The strength of Sveus is that the database contains real-time data that captures primary care as well as outpatient care. Data from the National Board of Health and Welfare contains, for example, no data from primary care. One weakness of the Sveus platform is that it, for legal reasons in the follow-up context, only contains data from public healthcare providers, which limits the sample for potentially relevant pilot studies. According to information received by TLV, there are legal uncertainties as to how data from private healthcare providers can be handled.

A platform that collects and harmonises complex data from different county councils (distributed data) like Sveus is an interesting solution. Such an arrangement makes it possible to compare outcomes between different county councils and to create a collective national follow-up at an aggregated level. Such a solution could potentially serve as a basis for not only ongoing follow-up of care in general, but also for drugs in clinical practice. However, it is not possible to evaluate the conditions as it is not clear how scalable the solution is, nor is it clear whether it can include all relevant and necessary data. The platform needs to be developed, and the legal conditions need to be clarified in order to meet TLV's follow-up needs. The more the county councils themselves use the platform, the better it can meet the needs of TLV in various follow-ups. Supplementary approaches will probably be needed in the future in order to be able to follow up drugs in clinical practice while awaiting a national optimal dataset being available for ongoing follow-up.

The full sub-report is available here: Pilot 4

2.5 Pilot 5: Data extraction via the national service platform

Pilot 5 aims to map out the possibility of retrieving data from medical records in one place, a type of "one-stop-shop". At present, separate data extractions are required from each system in order to retrieve data from different county councils with different medical records systems. Since many of today's medical records can

communicate their data in so-called service contracts, it should be possible to retrieve data from several different medical records systems from a single place with the same format. The service contract is used in the national service platform. This is what makes the national patient overview (NPÖ) possible. With NPÖ, a treating doctor can (with the patient's permission) gain access to relevant information about the patient's previous visits with other healthcare providers that may be outside of their own county council. The national service platform could be useful for building new cohorts and creating new knowledge by both collecting existing medical record data and adding separate questionnaires and other measurements. This requires that the patient gives their consent for data to be collected, the same as with other prospective research. The system could also be used to perform retroactive follow-ups of specific treatments at the individual level, where data is retrieved from the medical records. It is also necessary to look more closely at the legal conditions regarding what type of data can be extracted from the national service platform in a follow-up context.

The full sub-report is available here: Pilot 5

2.6 Pilot 6: Follow-up of cancer drugs

Many new cancer drugs have been introduced in recent years, and more are expected to come. Many of the treatment are costly, and the time for follow-up in the clinical trials is relatively short. Developed follow-up of the efficacy and safety of these treatments is therefore important. The purpose of Pilot 6 is to investigate to which extent data from a national and a regional quality registry can be used to support TLV's need for follow-up of medicines in clinical practice, from both a national and a regional perspective.

Overall methodology:

1) Choice of therapy area and drug group
2) Choice of drug and indication
3) Definition of relevant questions for evaluating the benefit and cost-effectiveness of selected cancer treatments and relevant variables
4) Investigation of whether the national registry for new cancer drugs can be used to answer the relevant questions, or if it can be adapted in order to answer the questions
5) Investigation of whether the regional registry for new cancer drugs can be used to answer the relevant questions, or if it can be adapted in order to answer the questions
6) Interview of the groups specified in 2.5 Action plan of the National Pharmaceutical Strategy, regarding their view of needs and opportunities in monitoring the efficacy and safety of the cancer drugs: namely SALAR [Swedish Association of Local Authorities and Regions], INCA [information network for cancer care], Region Stockholm, LIF [trade association for the research-based pharmaceutical industry]/relevant pharmaceutical company, and the Swedish Medical Products Agency.

7) Discussion of possible ways to improve the quality of data and simplify data extraction.

1) Choice of therapy area and drug group

The drug group PD1 inhibitors in the treatment of malignant melanoma was selected for the pilot study. PD1 inhibitors were selected because they are a relatively new drug category with both continuously expanded indications (areas of use) and increasing use. Very high levels of uncertainty were identified during TLV's health economic evaluations[4], so there is a need for supplementary analysis of benefit and cost-effectiveness. It is difficult to monitor the use and efficacy of PD1 inhibitors at present as they are used at clinics without the data being systematically registered. Thus, there is currently no access to complete information on the use of these treatments.

2) Choice of drug and indication

The indication *monotherapy in advanced (inoperable or metastatic) malignant melanoma* was selected since that indication has been available on the market for the longest period of time. Thus, that particular indication had the largest patient base for analysis. In malignant melanoma, efficacy is measured, among other things, in progression-free survival (PFS) and overall survival (OS). As a result, longer observations of data are required. Of the PD1 inhibitors currently approved, Opdivo and Keytruda have the indication in question.

The purpose of the pilot is not to compare the products with each other, but rather to examine the methodology regarding use of RWD to evaluate the efficacy and cost of the drugs.

3) Definition of relevant questions for evaluating the benefit and cost-effectiveness of selected cancer treatments and relevant variables

TLV mapped out what information would be needed to fully evaluate the cost-effectiveness of a cancer drug. Four groups with 16 different variables were identified as important in order to answer the evaluation questions:

- Patient characteristics
  - Age
  - Gender
  - Biomarkers
  - Date of diagnosis
  - Stage of disease
  - Performance status

- Treatment
  - Time of treatment start and end

[4] TLV's evaluation of Opdivo: Registration number: 4224/2014
TLV's evaluation of Keytruda: Registration number: 2100/2014

  - Treatment sequence
  - Dosage

- Outcome
  - Treatment response (e.g. complete response (CR), partial response (PR), no response (NR))
  - Survival
  - Progression-free survival
  - Reason for discontinuation of treatment
  - PROM/PREM measure, such as quality of life

- Cost
  - Cost of drug
  - Cost of care

4) Investigation of whether the national registry for new cancer drugs can be used to answer the relevant questions, or if it can be adapted in order to answer the questions (see "National part" below)
5) Investigation of whether the regional registry for new cancer drugs can be used to answer the relevant questions, or if it can be adapted in order to answer the questions (see "Regional part" below)
6) The interviews were conducted separately with all individual parties to discuss needs and opportunities related to cancer drug utilisation follow-up. All of the parties expressed that they saw a need for increased national cooperation between different parties, such as academia, authorities, companies and RCC, in order to improve opportunities for cancer drug utilisation follow-up.
7) Discussion of possible ways to improve the quality of data and simplify data extraction.

2.6.1 National part of the pilot study

In the national part of the pilot study, the purpose is to investigate whether the national quality registry for new cancer drugs ("Cancer drug registry") can be used to answer the relevant questions, or if it can be adapted in order to answer the questions.

Method

The national registry began registering data in 2018. Thus, it was not possible to perform a data extraction for the purpose of performing own analyses. A qualitative analysis focusing on future opportunities was therefore deemed best suited to answering the pilot study's question as to how the registry can be used to answer TLV's questions.

The method for the national part of the pilot was made up of a number of elements. In part, it involved comparing the existing variables in the registry with those

identified by TLV (presented above) in order to examine how many of them were available. It also involved a qualitative analysis of the registry report "Användning av nya cancerläkemedel – Redovisning gällande ett urval av cancerläkemedel med registrerad användning 1 januari – 30 juni 2018"[5] [Use of new cancer drugs – Report on a selection of cancer drugs with registered use 1 January – 30 June 2018], which was published in November 2018.

Data

As the registry is newly started, there is no established process for data extraction as of yet. A working group has been set up by RCC in collaboration to develop a clear work process for how to make data available. One possible way is to have a joint application procedure, where each steering group is responsible for approving extraction for its subset. TLV's interpretation is that such a process would simplify the application procedure, but that there is also a need for a procedure that shortens the lead time for obtaining data. For example, it is not a desirable situation for the time from approved application to actual data provision to be determined by the registry holder with the slowest process. It is therefore important that the working group addressing the issue develops a clear work process for how data can be made available without long lead times. Data from the registry will be useful for examining how a treatment is used across the country's regions for a specific indication. With the current variable composition, it is difficult to measure treatment effects in clinical practice using only data from the national registry. For this type of analysis, the data needs to be supplemented with relevant efficacy outcome measures and detailed patient characteristics from other data sources.

Results

The results section has been divided into three parts:

1. Which of TLV's questions for health economic evaluation can be answered with the registry? Which variables are available in the registry?
2. In what way can the registry be developed?
3. Qualitative analysis of the report "Användning av nya cancerläkemedel – Redovisning gällande ett urval av cancerläkemedel med registrerad användning 1 januari – 30 juni 2018"

<u>Which of TLV's questions for health economic evaluation can be answered with the registry? Which variables are available in the registry?</u>

The national quality registry contains data for a total of seven variables:

- *reporting hospital and clinic*
- *diagnosis*
- *performance status*

[5] https://www.cancercentrum.se/globalassets/vara-uppdrag/kunskapsstyrning/cancerlakemedel/kvalitetsregister/rapport_lakemedelsanvandning_6nov18_final.pdf

- current drug and combination therapy, where applicable
- treatment intention
- treatment period
- reason for discontinuation of treatment

The variable "Reason for discontinuation of treatment" contains the following alternatives: progression, toxicity, according to plan, other reason.

Data for these seven variables makes it possible to answer some of TLV's questions in a health economic analysis. It is of benefit to TLV that data from the registry have national coverage for follow-up of e.g. reimbursement restrictions or for analysis of equal access to care.

The variable "time to progression" is not included in the list. Access to information based on this variable would be of major benefit in TLV's work.

In what way can the registry be developed?
Not all of the variables that TLV identified as desirable are represented in the national quality registry. TLV therefore considers it desirable for the variables to be expanded over time.

Qualitative analysis of RCC's report "Användning av nya cancerläkemedel – Redovisning gällande ett urval av cancerläkemedel med registrerad användning 1 januari – 30 juni 2018":
The report presents information on treatment intention, performance status at the start of treatment, gender distribution and reason for discontinuation per drug. However, there is no information on what indication the results regard since the patient population was considered to be too small to be broken down by indication. Thus, it is not possible to read results specifically for malignant melanoma for Opdivo and Keytruda in the report.

There was reporting from all healthcare regions for a total of 1464 initiated treatments. Since diagnosis-specific patient overviews for lung, kidney and prostate cancer contains the same information to that found in the national cancer drug registry. Data from those registries was used to avoid double reporting.

There is no information on degree of coverage since the follow-up and registration time was short.

A report from the registry is scheduled to be published twice a year to provide access to national data on cancer drug utilisation.

Conclusions from the national part
It is of benefit to TLV for data from the registry to have national coverage for follow-up of e.g. reimbursement restrictions or for analysis of equality of care. It is an advantage that data can be transferred directly from the Stockholm registry and the patient overviews as these are compatible with the national registry. There is currently data for seven variables in the registry. Thus, some of TLV's need of data

on drug utilisation in clinical practice can be met. To be able to make full use of data from the registry in health economic assessments, TLV sees a need for the number of variables to be increased over time, and a structure for data extraction developed.

2.6.2 Regional part of the pilot study

In the regional part of the pilot study, the purpose it to investigate to what extent data from a regional quality registry can be used to answer TLV's questions (see above), or if it can be adapted in order to answer the questions. An additional aim is to examine whether data from the registries can be used to compare RWD with RCT.

Method

The method for the regional part of the pilot was made up of a number of elements. Two registries with data from the Stockholm region were selected as they were considered the most complete registries at present:

- RCC Stockholm-Gotland's Quality Registry for new cancer care drugs (Stockholm Registry)
- SLL's care database VAL

The existing variables in the registers would be compared with regard to the variables in the list created by TLV; see above.

Data from the registries would be combined to form the basis of a RWD cohort prior to a comparison between RCT and RWD for the drugs.

Data

The Ethical Review Board approved the study in July 2018. The research manager then sent an application for data extraction from VAL to the RUDE group [Council for External Data Disclosure] at SLL. Since the application had not yet been processed in December, data from VAL could not be included in the analysis. When data from the Stockholm registry was requested, there was initially uncertainty as to who needed to sign the personal data processing agreement, which delayed the process. Once the agreement was signed, the data was obtained relatively quickly. Experiences from the work with the pilot study illustrate that the policies for data disclosure need to be clarified and that lead times need to be shortened.

Results

The results section has been divided into four parts:

1. Which of TLV's questions for health economic evaluation can be answered with the registry? Which variables are available in the registry?
2. In what way can the registry be developed?
3. Analysis of data extraction
4. RWD-RCT comparison

<u>Which of TLV's questions for health economic evaluation can be answered with the registry? Which variables are available in the registry?</u>

The Stockholm registry contains data for a total of 9 variables that TLV needs in its work: age, gender, performance status, death, date of start and end of treatment,

dosage, treatment response and reason for discontinuation of treatment. Using these variables, TLV can answer a greater number of its questions than when using data available in the national registry.

The Stockholm registry also has the following variables registered, but there is not enough data for them to be included in the analysis: biomarkers, disease stage, treatment sequence and combination therapies. The variables in TLV's list that are completely missing from the Stockholm registry are: date of diagnosis, quality of life and cost of care.

In what way can the registry be developed?
As there was not data available for all 16 variables of TLV's list, TLV sees a need to increase the number of variables over time. The variable "time to progression" is not included in the list. Access to information based on this variable would be of major benefit in TLV's work.

Analysis of data extraction
Since only data from the Stockholm registry was available at compilation of this report, the analysis is based solely on the Stockholm registry's data.

In total, the Stockholm registry contained data from 168 patients treated with Keytruda or Opdivo. Data on patient characteristics was available for virtually all of the patients. The time for the start and end of treatment is also well documented. The dose information that is registered is "full or reduced dose", which means that both dose interval and dose size are missing. Time for survival is well documented in the database. Time to progression is registered as "reason for discontinuation of treatment". Since treatment can continue despite progression and treatment can be discontinued for other reasons, "discontinuation of treatment due to progression" does not correspond the same as the measure of progression-free survival often used in clinical trials. This means that the results of a data analysis of PFS cannot be compared with the results from clinical trials. Information for PROM/PREM (both of which are measures of quality of life) is missing, which makes it difficult to analyse quality of life.

RWD-RCT comparison
To be able to evaluate efficacy in RWD compared to efficacy in RCT, data from the two clinical trials was digitalised and the results in survival in the trials was compared with data from the Stockholm registry.[6] **Fel! Hittar inte referenskälla.** below illustrates available information for patient characteristics in the RCT studies and in RWD.

[6] Lumell's report "Uppföljning av cancerläkemedel med hälso- och sjukvård" [Follow-up of cancer drugs with healthcare]

Table 1. Available information for the patient characteristics in the RCT studies and RWD

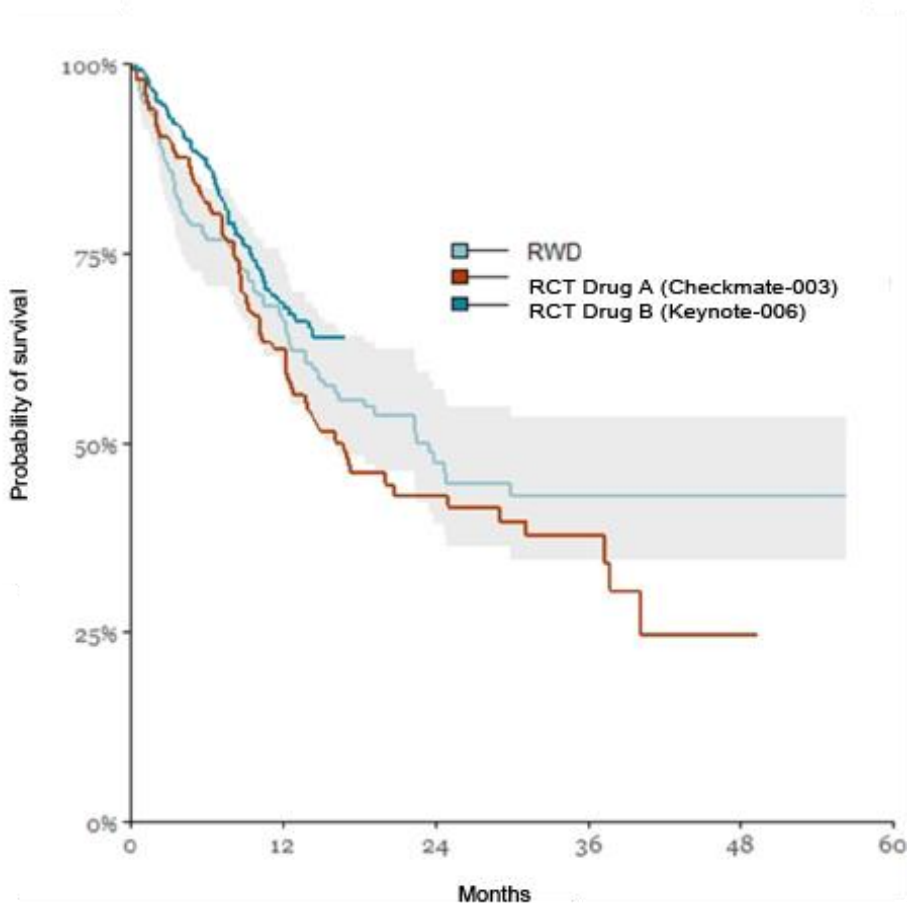
	Drug A (Checkmate-003)	Drug B (Keynote-006)	RWD Stockholm registry
Average age	61	63	63
Gender (entire study)	67 percent (72) men, 33 percent (35) women	58 percent (161) men, 42 percent (118) women	55 percent (93) men, 45 percent (75) women
Diagnosis	Metastatic melanoma	Stage III or metastatic melanoma	Metastatic melanoma
Patients in Kaplan-Meier population	107 patients	279 patients	168 patients
Performance status	ECOG (0–5): 0: 64 percent (68) 1: 34 percent (36) 2: 2 percent	ECOG (0–5): 0: 70 percent (195) 1: 30 percent (84)	WHO (0–5): 0: 77 percent (129) 1: 21 percent (35) 2: 2 percent (4) 3: 0

Figure 10 illustrates overall survival for RCT and RWD in the same figure. To avoid direct comparisons between RCT and RWD data for each drug, RWD from the two clinical trials was merged to form a common curve called RWD.

The survival results estimated via Kaplan-Meier curves look similar between RWD and RCT as the curves follow each other relatively well. However, due to the uncertainties, the results should be interpreted with caution, and no conclusions can been drawn regarding difference in efficacy between RWD and RCT:

- The information on patient characteristics was limited. Underlying differences between the patient population receiving Keytruda in clinical practice and the population receiving Opdivo in clinical practice can therefore not be ruled out.
- The analyses cannot be fully adjusted for underlying differences.
- The patients included in the RCT dataset for Opdivo had different doses of the drug: 0.1, 0.3, 1.0, 3.0 and 10.0 mg/kg. Thus, the data is not based solely on the currently approved dose of 3 mg/kg, which leads to uncertainty in the results.

Figure 10. Probability of survival in patients treated with Opdivo or Keytruda in the Stockholm quality registry (RWD) compared with clinical trials (RCT)



Conclusions from the regional part

There is currently a relevant degree of coverage for nine variables in the Stockholm registry. It thereby enables a greater number of TLV's questions in a health economic analysis to be answered than data from the national registry, which currently only has data for seven variables. To be able to make full use of data from the registry in health economic assessments, TLV sees a need for the number of variables to be increased over time.

Gaining access to the data was more time consuming than the time frame of the pilot study allowed. To be able to stay within the time frames, the analyses were based solely on data from the Stockholm registry. Data from the VAL database had not yet been delivered at the pilot time's cut-off point. As a result, analyses of cost of care could not be carried out.

The survival results estimated via Kaplan-Meier curves look similar between RWD and RCT as they follow each other relatively well. However, due to the uncertainties in the material, the results should be interpreted with caution, and no conclusions can be drawn regarding difference in efficacy between RWD and RCT. The pilot study should primarily be considered a step in the continued development of methods for using RWD. The full sub-report is available here: Pilot 6

3 Lessons learned in relation to data access and methods

3.1 Introduction

Access to relevant data and robust analysis methods is key to all drug utilisation follow-up. What opportunities for follow-up exist can affect TLV's price and reimbursement decisions, especially during the introduction of a new drug. When a new drug is introduced, there is often uncertainty regarding use and efficacy in clinical practice. It may therefore be necessary to initially limit reimbursement to the patients with the greatest medical need. This is especially true for expensive drugs. At a later stage, once more knowledge has been generated, reimbursement can be expanded to also include patients with lower medical needs. Alternatively, if knowledge shows that the drug is less effective than expected, the cost may need to be lowered for reimbursement to continue, or use may need to be phased out. The ability to follow up drug utilisation and treatment effect in clinical practice can therefore accelerate the introduction of new drugs. Knowing that knowledge will be generated that can be used to adjust how a drug is used in practice can lead to a higher degree of acceptance of the uncertainties associated with introduction of a new treatment, both at TLV and by prescribers.

Access to relevant data and robust analysis methods can help patients gain faster access to new, innovative drugs at a cost that is reasonable for society. Knowledge about drug utilisation is also key to care being steered towards more knowledge-based, equal, and economically sustainable care of high quality.

3.2 Experiences related to methodology

The example of Pilot 1 (PCSK9 inhibitor) clearly illustrates the complexity of interpreting clinical data. One the one hand, it is easy to extract data to be able to perform an evaluation. At the same time, it is challenging to interpret the data correctly. At first glance, Repatha and Praluent seem to differ in efficacy as measured in terms of probability of cardiovascular event or death. When you stratify the data and compare Repatha and Praluent at the same time point, it is clear how similar they are. The fact that patients were put on Repatha early after the decision on reimbursement is most likely the result of sicker patients, who are more likely to have a cardiovascular event or die, being put on the drug initially. Once the treatment has been available for some time, use spreads to patients who are not as sick. Naturally, because they are in slightly better health, they have a lower probability of suffering a cardiovascular event or death. Thus, the different groups cannot be compared with one another without first correcting for differences in health.

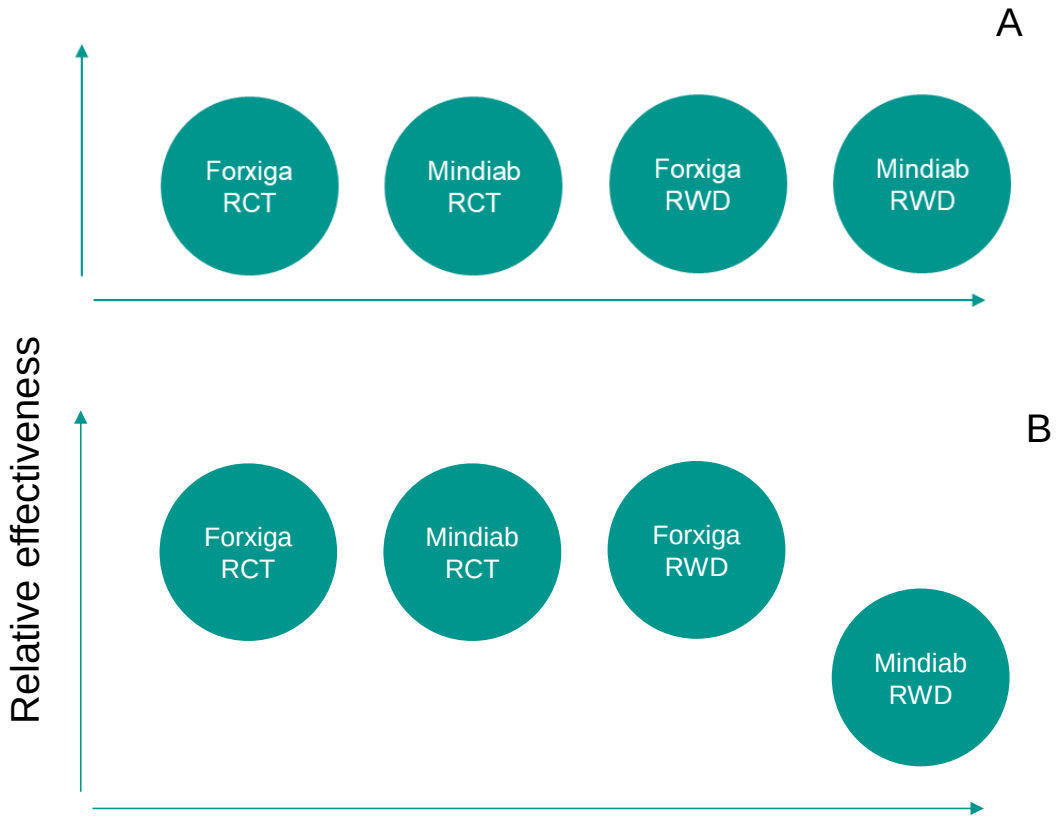
When you then compare the likelihood of having an event based on RWD, it may appear that the efficacy of the drug is much worse in reality than it is in the clinical trial. But, even patients treated with placebo in the clinical trial have a much lower likelihood of a cardiovascular event than patients with PCSK9 inhibitor in clinical practice. So, the explanation is more likely that the patient population receiving the drug in clinical practice has different characteristics than the one studied in the trial. Thus, comparing efficacy in clinical trial and efficacy in clinical practice risks becoming a classic example of comparing apples and oranges. Repatha and Praluent have reimbursement restrictions, which means that patients treated with PCSK9 inhibitors in clinical practice are sicker than the patients who were included in the clinical trial. If you compare Repatha and Praluent in the same time period, it appears that efficacy is about the same.

However, the description needs to be further problematised. If one of the two drugs is given to slightly sicker patients, who are more likely to have a cardiovascular event, but this drug is slightly more effective, then the efficacy of the two drugs may appear equal even though they are not. More high-resolution data is needed to be able to describe differences in relevant patient characteristics. In the example above, only the time dimension could be taken into account.

The same could apply to what we see in the example with PD1 inhibitor (Pilot 6). The survival results estimated via Kaplan-Meier curves look similar when comparing RWD and RCT as the curves follow each other relatively well. The key question is whether the patient populations are really the same as in the trials. Or could we have apples and oranges in the different baskets? To answer this, we need information about both the individuals in the clinical trials and the individuals using the drugs in clinical practice. Since we do not have access to this type of data (neither from the quality registry, nor from the clinical trials), we cannot say whether the patient populations in clinical practice are the same as in the trials.

In the example with diabetes (Pilot 2), an attempt is therefore made to deal with the problem of apples and oranges by weighting the patient population from the clinical trials with the registry data. By having access to efficacy data in both the control group and the active group in the clinical trial as well as in clinical practice, four different groups were created to compare with each other. Bridging RCT and RWD with each other makes it possible to examine how the efficacy observed in clinical trial bridges to clinical practice. In the example with Victoza both the study drug and the control group bridge well to RWD when you examine body weight. In contrast, HbA1c scales poorly for the Mindiab control group to RWD, despite available patient factors being used. This can be illustrated in Figure 11, where A describes how e.g. body weight scales while B describes how HbA1c scales.

Figure 11. Illustration of how relative efficacy scales in the diabetes example, where A is the scaling of an outcome measure that scales well between the different groups, while B represents poor scaling.



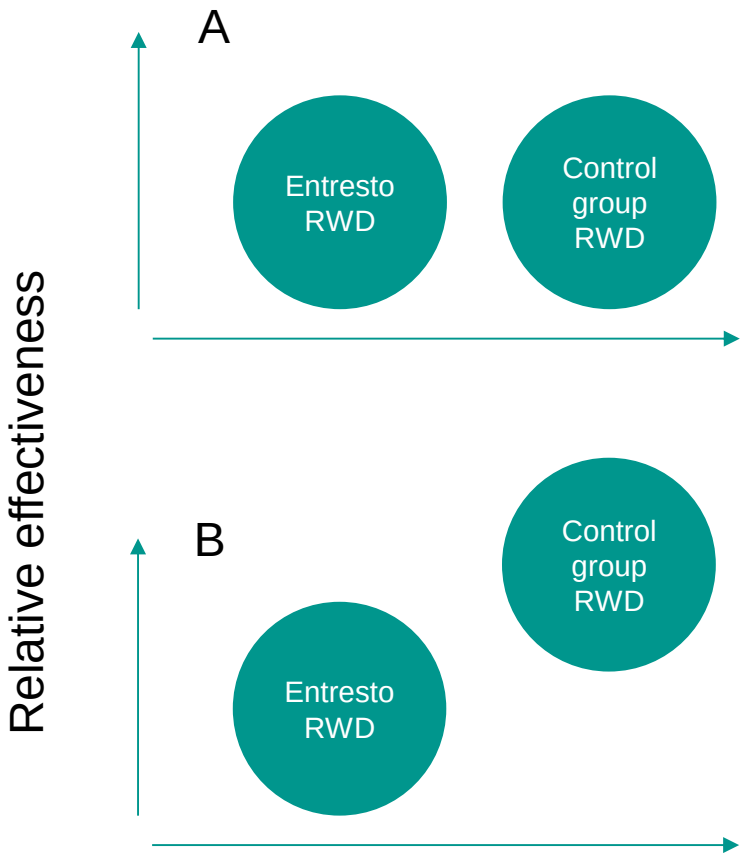
Patient characteristics are relatively similar in RCT and RWD. The application of scaling factors has no major impact on the estimated efficacy and thus cannot explain why patients treated with Mindiab have differences in clinical practice.

If RCT data is not available or the patient population differs between clinical practice and RCT, the evaluation must be done in a different way. In the example with Entresto (Pilot 3), a control group was created via medical record data. Extensive work was done prior to data extraction in order to identify/define factors that could affect either choice of treatment and/or the efficacy of treatment. In the clinical trial on which the Entresto reimbursement decision was based, the control group was ACE inhibitors. In the preliminary work, four different treatments (including pacemaker) were identified as comparable to Entresto in clinical practice. In the earlier example with PCSK9 inhibitors, a clear drift over time was seen in the likelihood of a cardiovascular event. However, the same phenomenon could not be seen in this example. The evaluation showed that treatment with Entresto seems to work just as well as for the comparison group. When compared to published RCT data for Entresto, the likelihood of a cardiovascular event is higher in clinical practice, which confirms that the patient population receiving the drug in clinical practice differs from the one included in the clinical trial.

Many different factors are included in the medical record data, and these can be used to match patients or weight the results. The inherent problem with this type of

analysis is that you can only apply weighting for information you are aware of and have access to. In the Entresto example, it looks like the variables used as controls produce comparable groups. But, how do we know that both groups are apples? An attempt to visualise the problem is found in Figure 12.

Figure 12. Visualisation of relative efficacy in the Entresto example



When we compare Entresto with the control group in A, the two groups seem to be comparable, but what would it look like if we included additional explanatory variables that are not available in the medical record data? Would example A still be valid, or would we slip over to B in Figure 12. How do we ensure that this is not the case, so we that the results can be used for decision-making? In the case of Forxiga in RCT and RWD, we have multiple groups to compare with each other. We can then see that one of the groups does not scale well. But, what do we do when we cannot do this exercise with multiple groups, but instead only have two groups? One way forward could be to find several different ways to verify that the control group is actually comparable with the treatment group you want to evaluate. If several independent analyses point in the same direction, there is probably a greater inclination to rely on the results when they are to be used for decision-making. There is therefore a need to identify alternative methods for verifying results.

3.3 Experiences related to data access

TLV's work with the pilot studies within the framework of the two Government assignments shows that there is great potential and a great need to develop follow-up of drug utilisation and treatment effect in clinical practice. An important conclusion is that the current system houses great knowledge in the form of extensive data that could be used for drug utilisation follow-up. However, change and continued development are needed to transform the potential into actual relevant follow-up.

3.3.1 Insufficient data inhibits innovation

TLV's experience shows that there is relevant data to better monitor drug utilisation and treatment effect in clinical practice. This is a key prerequisite for TLV to be able to achieve its goals of developing value-based pricing and promoting innovation by promoting use of new, innovative drugs while ensuring a reasonable cost to the public. As things stand now, TLV finds data access to be insufficient to be able to fully evaluate the methods and thereby the results. The shortcomings are, in part, that it is not possible to produce all relevant variables in a structured and comprehensive way (e.g. progression-free survival in Pilot 6). The information that TLV needs to access is also spread across different data sources, making it difficult to access relevant datasets.

Unclear processes and long lead times related to data disclosure were a common element of all of the pilot studies. The processes differ from one county council/region to another, as well as between public and private actors. Pilot 3 shows that it is labour-intensive to gain access to medical record data from a county council. If there is no change in the possibility of data access, very extensive efforts will be necessary to scale this up to a national perspective.

In the current research environment there is no actor who needs, or rather demands, data at a detailed level and in (practically) real time. Although the content does not differ significantly, TLV often has different needs for access to data than researchers do. TLV's follow-up must be carried out in close connection with the introduction of a new treatment in order to be valuable and be used to develop value-based pricing and promote the use of innovative drugs.

Even within the framework of decisions regarding new drugs, which may take no longer than 180 days, data linked to existing treatments is needed e.g. to validate the results in health economic models. The current structures do not meet TLV's needs and therefore require development. TLV is not alone in its requests for detailed follow-up. County councils also have such needs within the context of new therapy introduction. A trend towards some kind of "one-stop-shop" with one entry point to multiple data sources would significantly support a more continuous and agile monitoring and thereby promote early access to new drugs.

In the work with Pilot 6 (follow-up of cancer), interviews were conducted with relevant actors (SALAR, INCA, SLL, LIF/relevant pharmaceutical company, the

Swedish Medical Products Agency, etc.). All of the interviewees saw the need for stronger coordination at the national level.

TLV therefore needs to continue developing work methods and forms of collaboration with relevant actors to improve the Agency's access to relevant information. Improving access to data is key to not only verifying the use of the analysis methods tested within the framework of the current Government assignments but also testing other methods. A wide range of analysis methods is required to be able to handle the complexity of (without randomising) creating a comparison group that is suitable for evaluating treatment effect in clinical practice. To be able to obtain robust conclusions that can also be used as the basis for decision-making, it may be necessary to use different methods in parallel to investigate the same question.

3.3.2 Challenges and opportunities

Even when there is relevant data that can be structured, it is a challenge to gain access to the data in a timely manner since it is often spread across different sources, each with different routes to decision-making. Datasets from different sources often need to be linked together, which contributes to long lead times. For example, data may be needed from various registries that are managed by authorities such as the National Board of Health and Welfare or Statistics Sweden, or from respective quality registries and regional registry centre organisations.

A separate application must be made to the individual managers, each of which has different processes and different capacities. This affects lead times and which data is disclosed. The organisation with the slowest process will determine when the work can start. Access is further complicated in cases where the required data comes from medical records systems. This is due, in part, to the application having to be made to a larger number of principals, often within the same county council, as well as ambiguities in the data disclosure process.

In the Entresto evaluation (Pilot 3) it took individual interaction with seven different healthcare providers *within* a single county council to gain access to data. In many cases, the county councils and regions do not have sufficient capacity to provide data in a way that meets TLV's needs. In addition, data from medical records systems must be processed before it can be used adequately. This type of processing usually has to be adapted and tailored for different county councils as either the same medical records system or a code system variant is used nationally. Moreover, different actors may interpret legislation differently, which means that the possibility of disclosing data may be assessed differently even though the same type of information is being used. In some of pilot studies, it took more than a year from project start to receipt of data. As a result, analyses could only be performed for a few months. Some data was even delivered after the cut-off point for inclusion in the analysis.

A key challenge is to find infrastructures that effectively extract data from different systems, and harmonise and structure this data. Another challenge is to make data

more readily available to relevant actors, such as academia, authorities and companies. There is a need for a clearer and more transparent data disclosure process. Above all, the processes need to be coordinated and the lead times shortened.

A developed and in-depth collaboration with quality registries and registry managers is required. Quality registries possess a lot of the expertise required to be able to use and interpret the complexity in the data. In many cases, it is the quality registries that have a large number of the relevant variables, even though they do not provide full coverage.

Access to data is key to follow-up. Exactly which data and at what level depends on the question being investigated. In some cases, all that is needed is simpler statistics at the anonymised individual level from e.g. the National Board of Health and Welfare's drug registry. Other cases may require more complex data with high resolution in terms of time, variables and geography. This type of complex data is rarely available in a single data source, but may come from e.g. patient records, quality registries and health data registries that must be linked together. Situations where not all of the data requested is available is more the rule than the exception. A basic prerequisite is that data management follows strict requirements for patient privacy and security protection.

Data need is governed by the question being investigated. Two examples are presented below, one where scaled data is used (Pilot 1) and one where detailed data is used (Pilot 3).

The purpose of Pilot 1 was to investigate the level of compliance with the reimbursement restriction for PCSK9 inhibitor in clinical practice, and to gain more knowledge about the recurrence of cardiovascular events and/or death for individuals treated with these drugs. The starting point was data that TLV ordered from the Swedeheart quality registry and data from the National Board of Health and Welfare's drug registry and patient registry.

The data order from Swedeheart was simpler in nature and consisted of LDL level *before* and *after* initiation of PCSK9 inhibitor in patients who recently suffered myocardial infarction. Of those included in Swedeheart, only a small number of individuals had been treated with PCSK9 inhibitor. It was therefore not possible to draw any conclusions regarding compliance or non-compliance with the reimbursement restriction. The registry did not capture the patient population relevant to the follow-up, and no other quality registry with better data is available.

In the same pilot study, prescription data from all dispensings of PCSK9 inhibitor was used. Data showed that patients in clinical practice are more likely to suffer a cardiovascular event than individuals in the clinical trial, probably because they are generally sicker. As TLV does not have access to the National Board of Health and Welfare's data, there was limited ability to obtain a dataset with individual information that could aid in understanding what differentiates patients in clinical practice from the patients in the clinical trial. If TLV were to be able to process more

high-resolution data from the National Board of Health and Welfare's health data registry, the analyses could provide a more nuanced picture of drug utilisation. This does not just apply to registry data from the National Board of Health and Welfare, but can also be extended to socioeconomic conditions from e.g. SCB and sick leave history from Försäkringskassan [Swedish Social Insurance Agency]. In certain situations, such analyses would provide a sufficient knowledge base. In other situations, the analyses could lead to determining what type of medical record data may be relevant to use for further evaluation.

In Pilot 3, the aim was to evaluate the treatment effect of Entresto in clinical practice by creating a comparison group in clinical practice. In terms of data, the challenge was gaining access to information about the patient populations that was detailed enough to create two groups that are completely equivalent except for which treatment they are receiving. The comparative alternative for Entresto in clinical practice was determined to be the highest tolerable dose of beta blocker, an ACE inhibitor/ARB, aldosterone antagonist plus add-on therapy in the form of either ivabradine, digoxin or ICD/CRT. In order to create comparable groups, the analyses have to be adjusted for a number of factors that could otherwise throw off the results.

This type of analysis requires detailed information about individual characteristics and a number of clinical variables, which necessitates access to medical record data. During data processing work, it came to light that some relevant information needed to be able to select part of the comparison group was missing from the medical records. The analysis plan was therefore adjusted based on data access, despite the fact that it was the best individual data available. The results of the analysis cannot demonstrate that Entresto has better efficacy than the treatment in the comparison group described above. It is worth noting that the comparative alternative in Swedish clinical practice is different than the comparative alternative in the clinical trial, where Entresto was compared with an ACE inhibitor. The method has great potential, but both the results and the method require verification. Method verification could be done by applying the method to other drugs or to data from other county councils, while the results could be verified by applying other methods to the same question.

The example with diabetes (Pilot 2) shows both difficulties and opportunities from using supplementary data from clinical trials. It is interesting that one of the pilot studies focuses on drug treatments where the patient populations in RCT and RWD are very similar, while two other pilot studies (Entresto and PCSK9 inhibitor) focus on areas where they are not similar. In order to further evaluate the possibility of using data that the companies generate, the infrastructure for data sharing must be improved. Clear rules and procedures for the conditions under which data may be disclosed to a third party are a necessity in order to build trust between the Agency and industry. The external platform used in the pilot study is a possibility, but was difficult to work with in practice. Alternative solutions should be explored in order to fully integrate RCT and RWD data with each other. The legal conditions in particular must be investigated.

The national new cancer drug registry contains information on, inter alia, duration of treatment and individual characteristics such as age and gender, broken down by county council for a selection of new cancer drugs. Although the information is limited, the national part of Pilot 6 illustrates that the registry provides a valuable opportunity to monitor the introduction of cancer drugs at the national level. The registry can provide an indication of whether utilisation is equal across the country. There may, of course, be demographic, genetic and socioeconomic variations between county councils that the analyses may need to correct for to make it possible to draw conclusions as to whether utilisation is equal or not. There may also be regional differences in treatment practice, which may mean that the medical need of a specific drug may differ between county councils. For example, the treatment of choice for a specific medical condition may be surgery in one county council, while treatment with drugs is more common in another. In order to perform a complete analysis, information is also needed about the individuals who do not receive the drug in question. Only then can a full needs analysis be performed.

To get a more nuanced picture of drug utilisation, a larger number of variables need to be analysed. The regional part of Pilot 6 illustrates how an expanded set of variables can be used to create an in-depth follow-up of the introduction of Opdivo and Keytruda. For individuals belonging to SLL/Gotland, the variables were expanded with information on, inter alia, duration of treatment and survival. This made it possible to also study the survival of the individuals treated with the drugs in question. It takes a larger number of variables describing factors such as patient characteristics and disease severity in order to estimate treatment effect in clinical practice. The outcome measure PFS (progression-free survival) would also need to be measured in the same way in both the clinical trial and clinical practice in order to be comparable. Moreover, a reference population is also needed from either clinical practice or the clinical trial. In the latter case, individual data from the clinical trial is required.

4 Continued work

TLV finds that continued work on the development of real world data (RWD) is a key prerequisite for achieving the goals of early and equal access to new, innovative treatments at a reasonable and sustainable cost for the public. With this in mind, TLV is requesting continued Government assignments in order to create clear and stable conditions for continued development work.

During the course of the work, it has become clear that there is great interest in these issues and this type of development. Authorities, academic institutions, registry holders, healthcare providers and a large number of private actors (both pharmaceutical companies and data analytics companies) have all expressed a willingness to get involved. There is considerable development potential in this area, and some of the challenges relate to developing collaborative platforms in terms of both knowledge and funding. The great potential of data generated in Sweden coupled with knowledge and the positive commitment is something that Sweden should take advantage of and develop as a life science nation.

TLV's ultimate ambition is to be able to use follow-ups based on data from clinical practice as a basis for decision-making. Such an ability could accelerate the adoption of new drugs and thereby enable new, innovative drugs to reach patients more quickly. Development on multiple levels is needed to achieve this.

The methods evaluated in the pilot studies have advantages and disadvantages. Further work must be done to evaluate how and under what conditions these methods can be used as a basis for decision-making. Since it is likely that different methods will have varying degrees of success in different contexts, there is a need to evaluate multiple methods. The results of a method also need to be verified using different approaches in order to evaluate their robustness. Procedures need to be worked out to be able to initiate a follow-up more quickly and to better capture data that is relevant for answering the question. It is the question that should steer which data should be used rather than which data can be accessed steering which follow-ups are performed.

As previously stated, TLV will have a great need to collaborate with multiple actors, such as academia and private actors, who are contracted to develop and evaluate methods. One experience from the assignments is that TLV needs to work in close collaboration with the contractors. Because the analyses continuously provide knowledge that can, in turn, affect the question, TLV needs to be closely involved to make it possible for the results of a follow-up to be used in the Agency's activities. It is difficult for an external party to be able to take in all of the perspectives that TLV needs to relate to, and it is not always possible to specify these in advance before knowledge about utilisation and efficacy is in place. Collaboration with other actors should be facilitated if TLV was given the capacity to handle individual data. The

lack of such data currently limits what analyses TLV can perform internally. Changes in legislation are required in order to gain access to individual data.

Quality registries remain an important source of data, but the work that has been done shows that data from these registries may need to be supplemented with medical record data. Other approaches need to be tested in order to capture data that has a larger number of variables, greater coverage, and contains up-to-date information. One of the pilot studies indicates that the national service platform may be a potential source of data. Several quality registries are already running through the platform for automatic update of registry content. Another pilot study shows that it is possible to link the county councils' administrative systems with information about drugs and visualise the data on the Sveus analysis platform. There are also private actors who work directly with medial record data from the county councils in different contexts.

Data from the National Board of Health and Welfare is an important foundation. The potential of data from the National Board of Health and Welfare has been enhanced by the fact that data sent to the patient registries is now registered on a running basis throughout the year. A shortcoming in the patient registry is that it is currently not mandatory for primary care to report data. There is a common need to find ways to adapt the data extracted to ensure it is relevant, structured and analysable, and that it can be used to develop methods for evaluating or monitoring drug effects.

5 Appendix

The sub-reports for the various pilot projects are available for download via the links below.

Pilot 2

https://www.tlv.se/download/18.3d61c3c3167c695651ef749/1545299007346/Pilot2-Metod_for_utvardering_av_behandlingseffekter_i_klinisk_vardag_med_en_jamforelsegrupp_fran_en_klinisk_provning.pdf

https://www.tlv.se/download/18.3d61c3c3167c695651ef748/1545299007295/Pilot2-Appendix_inklusionskriterier.pdf

Pilot 3
https://www.tlv.se/download/18.3d61c3c3167c695651ef74a/1545299007466/Pilot3-Metod_for_utvardering_av_behandlingseffekter_i_klinisk_vardag_med_en_jamforelsegrupp_fran_klinisk_vardag.pdf

Pilot 4
https://www.tlv.se/download/18.3d61c3c3167c695651ef74b/1545299007557/Pilot4-Lakemedelsuppfoljning_pa_SVEUS-analysplattform.pdf

Pilot 5
https://www.tlv.se/download/18.3d61c3c3167c695651ef74c/1545299007618/Pilot5-Datauttag_via_nationella_tjansteplattformen.pdf

Pilot 6
https://www.tlv.se/download/18.3d61c3c3167c695651ef74d/1545299007680/Pilot6-Uppfoljning_av_cancerlakemedel.pdf