University of Lund

Department of Economics

Master's Thesis in Economics

Marcus Brafors

Supervised by Carl Hampus Lyttkens

2016-07-29

On the evaluation of pre-hospital telemedicine in acute stroke care

Keywords: Telestroke, Telemedicine, Stroke care intervention, Economic evaluation, Cost-Effectiveness analysis.

Abstract

Purpose: To propose a blueprint of how the policymakers responsible for health care reforms in the county of Scania (Sweden) could evaluate a prospective stroke care intervention from a societal point of view. More specifically, the intervention will consist in the introduction of telemedicine technology to their ambulance fleet. In addition, key methodological issues, of interest to the policymaker, are to be discerned and commented on.

Method: The vantage point will be a Markov simulation. The various methodological considerations pertaining to such simulations will be highlighted. In particular, the variables one ought to consider will be thoroughly discussed. Furthermore, there will be suggestions as to what kind of data the policymaker will need in order to make the simulation as interesting and realistic as possible. The definition of the outcome variables will be wide so as to encompass a societal point of view.

Results: At the moment, the direct (and indirect) costs associated with the intervention seem rather uncharted. In addition, there are no studies indicating that there indeed is a clinical effect of the intervention. A simple way to facilitate the evaluation would be to try and keep track of (i) how discharged stroke patients are distributed across mRS-categories, and (ii) how the intervention affects this distribution. The main reason for this is that a lot of academic effort has already been invested in relating mRS to health care costs, quality of life, and mortality.

Contents

1. Introduction and outline

Most of the major changes in the Swedish health care system are somehow administered (or at least dictated) by the policymakers in the county council. This will be true also of a forthcoming (albeit prospective) reform of the stroke care in the county of Scania (Sweden). As will be described more in depth in subsequent sections, stroke is a widespread condition. What is more, in most cases it will without fail seriously disable the patient and consequently reduce the patient's quality of life. Hoping to mitigate this negative impact, a telemedical intervention in the stroke care has been suggested. More specifically, the ambulances deployed in the acute phase could perhaps be equipped with cameras so as to allow the ambulance personnel to communicate with a team of stoke experts that are, as it were, less mobile. These experts could then contribute with a preliminary diagnosis that would make the post-acute phase of the treatment more efficient and less cumbersome.

However, before carrying out this reform, it will of course be beneficial to consider how to properly evaluate it once it has been entrenched. Performing economic evaluations of health care reforms will allow the policymakers to be confident in its presumed effects. In addition, the stakeholders will be able to hold the policymakers accountable in case the reform is somehow not satisfactory. In fact, it is these latter two considerations that make up the rationale for the present thesis.

The reader will not be reading a complete and full-blown evaluation of the aforementioned intervention (mainly because the intervention in question has yet to be implemented). Rather, the present thesis should be perceived as something of a blueprint. It will suggest how an economic evaluation of this stroke care intervention could be carried out. It will hold suggestions as to what methods should be used and it will comment on what data sources are (and are not) presently available. It will also carefully articulate whatever methodological issues the economist (as well as the policymakers) will be forced to confront when carrying out the evaluation.

To begin with, there will be a brief sketch of the medical details of the stroke condition. There will also be a short survey of the impact stroke has on society. The details of the stroke care intervention considered by the county will then be presented. Admittedly, at the time of writing these details are not as rich as one could have hoped for but they will definitely suffice to give the general idea. All of the above will take place in section **2.1**.

The subsequent section, **2.2**, will hold some thoughts on the evaluation of health care

intervention in general and in particular it will give the details on how to create a so-called Markov model. In addition, it will introduce the concept of the modified Rankin Scale as a measure of disability and indicate how it can be used in a Markov context.

Each of the variables, hinted at in chapter **2**, will now be devoted a chapter each. First off is the QALY variable. There, one will find discussions of methodological topics of interest to the policymaker. These topics include the distinctions between intermediate and final measures of health as well as that between disease-specific and generic measures. There will also be comments on the particulars of eliciting QALY weights later to be used in the Markov simulation aforementioned. This will all take place in chapter **3**.

Chapter **4** will hold details on the mortality variable. It will be perhaps the briefest of all the chapters. There will be a suggestion as to how one constructs transition probabilities to the absorbing state in the Markov model. The chapter will also contain some comments on whether these probabilities should be sensitive to the change of cycles in the model.

Then, chapter **5** will hold comments on the various components of the cost variable. These components include direct and indirect costs as well as their respective sub categories. Direct costs will be seen as being rather unproblematic while the indirect ones will not. The latter kind of cost will be divided into so-called unrelated costs and productivity cost. Both of these components will prove to be problematic to say the least.

Lastly, chapter **6** will round off by summarizing the main points of the preceding chapters as well as contribute with concluding remarks.

2. Preliminaries

First, a brief sketch of the medical background in **2.1**. The consecutive section **2.2** will suggest how to evaluate a stroke care intervention. This suggestion will include the so-called Markov model as well as the modified Ranking Scale (mRS). Lastly, previous research will be commented on in section **2.3**.

2.1 Medical background

This sketch includes the health impact of stroke on society as well as how the health care system attempts to brace this impact. One will also find a description of the intervention that is being considered by the county.

2.1.1 The condition

Stroke is the clinical diagnosis denoting symptoms of brain function disturbances that enhances at a quick pace and that is presumed to have a vascular cause. Also, for the term to be used in its proper sense, this condition must either result in death or last for at least 24 hours. The symptoms include difficulties with speech, weakness throughout arms and legs as well as face drooping. Characteristically, they often afflict but one side of the body.

In case there is a lack of continuous blood flow to the brain it will be deprived of nutrients and oxygen. This deprivation, in turn, will cause immediate damage. The disturbances in blood flow will stem from either of two sources. Ischemic strokes are the ones resulting from a blockage while hemorrhagic ones result from a bleeding (Berglund et al., 2006). As for risk factors, some of the more salient ones are age, physical inactivity, smoking, diabetes, and hypertension (Ghatnekar, 2013).

It has been estimated that stroke brought approximately 6 million casualties worldwide in 2008. The incidence in Sweden amounts to 25,000 annually although the number of casualties has fallen the last 15 years (Ghatnekar, 2013). As for the survivors, most will suffer from severe disability, mental as well as physical, the rehabilitation of which is both outdrawn and demanding. Obviously this is detrimental to the quality of life of the patient. In addition, the quality of life of the patient's caregivers is also negatively affected (Olai, 2010).

In 2014, the share of male stroke patients (in Sweden) was 52% while 48% were female. These proportions have changed little over the years. The same is true of the average age which is

about 75 although slightly lower for males (73) and slightly higher for females (78) (Norrving, 2016). Since the acute phase of the condition is often followed by severe impairments, a considerable amount of resources are devoted to secondary prevention and rehabilitation (including physiotherapy as well as speech therapy). Many patients will also require support with so-called "activities of daily life" (ADL) for the remainder of their lives. In Sweden, these health care services are provided by the municipality (although, of course, a lot of responsibility falls on informal caregivers)(Norrving, 2016).

2.1.2 The intervention

Once the patient arrives at the emergency room (ER), there are ways to restore the flow of blood to the brain (thrombolysis, thrombectomy). However, even though such an incision might save the patient's life, there could still be severe complications after the acute phase. The extent of these complications will be greater in case the stroke is not alleviated straight away. Put differently, the longer the flow of blood to the brain is disturbed the more severe will the post-acute complications be.

In case of an alleged stroke, the patient's symptoms (such as difficulties with speech, face drooping and so on) will be considered. Then, given that stroke seems a likely diagnosis, the patient will be taken to the ER. If deemed necessary, the patient will then undergo a CT scan (sometimes at a different medical facility) so as to make a final verdict possible. Once all this information has been gathered the most appropriate treatment will be decided on. The "visual" symptoms can be examined by the ambulance personnel to some extent (so as to give a preliminary diagnosis) but the final verdict must be made by a neurologist[1].

The intervention consists in the installation of two cameras on the inside of the ambulance. By means of these cameras, vision and sound, describing the patient's symptoms, will be transmitted from the ambulance to the on-call stroke team. This will be a two-way live transmission and the cameras will be controlled by the ambulance personnel. The contention is that this intervention will facilitate the pre-hospital stroke care in the following ways:

- Increased precision in making pre-ER judgements concerning which patients should be sent directly to facilities holding a CT scanner (and thus avoiding the time consuming detours to the ER).

[1] The particulars of the intervention as here described comes from private correspondence with Marie Grey (Region Skåne).

- Increased precision in making judgements as to what patients have been recognized too late in order for thrombolysis to be feasible. Instead such patients can be directed elsewhere.
- Quicker identification of patients eligible for thrombectomy. Because of this, ambulance personnel will remain at the CT-scan facility, ready to transport the patient on to the next facility where the thrombectomy will take place. In addition, the medical staff at this second facility will be notified much earlier and thus be better prepared.

Equipping ambulances with this technology, then, is expected to cut the duration of the pre-treatment phase. Put differently, one hopes to shorten the time elapsing from the onset of symptoms to the initiation of the proper treatment.

2.2 Evaluating stroke care interventions

First, the Markov model will be introduced in **2.2.1**. In particular, this introduction will elaborate on transition probabilities, the choice of time horizon, but also discount rates. The subsequent section **2.2.2** will describe the modified Rankin Scale and also suggest how this concept can work together with the Markov model.

2.2.1 The Markov model

A convenient way of charting how a certain disease will affect a cohort of patients is to implement a Markov model[2]. In order to do this, one must first designate the so-called Markov states. These are simply health states that are closely associated with the disease in question. In our case, one of these will be the state of death. This state will eventually, as it were, absorb all the patients in the cohort. Typically, the remaining states can be ordered from "better" to "worse" so as to allow the (simulated) patients to degenerate towards the absorbing state. In the case of stroke, it will make sense to have the Markov states represent disability of varying severity.

Once the Markov states are laid out, they will be related to one another by means of so-called transition probabilities. These will specify, for each pair of states, how likely it is that a patient will transition from one to the other. If one is presently in state A, say, one might transition into state B with probability p_1 or, perhaps, into state C with probability p_2. However, the transitions will only take place once every cycle. These cycles could represent months or years

[2] For the description of the Markov model this section relies heavily on Drummond et al. (2015).

or whatever chunk of time fits the disease under consideration. For our purposes, since the states will represent different severities of disability, one could think of the transition probabilities as risks of relapsing into a new stroke (resulting in perhaps more severe disability).

It is important to note that transition probabilities, such as the risk of relapsing into a second (or perhaps third) stroke, should, in some instances, be made into functions of the current cycle. For example, the risk of dying will increase with age (the mortality of a 20-year-old is much lower than that of a 75-year-old). This point becomes clearer if one considers the fact that the risk of stroke increases with age. Obviously this suggests that, as the cohort moves down through the cycles, the risk of relapsing will have to increase. The same is true of the probability of transitioning into the absorbing state. In other words, the more strokes you have suffered the less likely are you to survive yet another stroke (Ghatnekar, 2013).

The time horizon of one's simulation is given by the number of cycles. It will often make most sense to have a time horizon that is as long as the average remaining life span of the patients in the cohort. The reason for this is that many health care interventions have repercussions throughout of the remainder of the patient's life. A case of stroke, for example, will cause impairment. If the severity of this impairment could somehow be mitigated by an intervention, the patients will benefit for the rest of their lives. Less severe impairment will also imply less costly rehabilitation. The only way to (at least approximately) account for these life-long benefits and costs is to adapt a life-long time horizon. In terms of the model, the end of the time horizon should then be a cycle where a majority of the cohort is in the absorbing state (e.g. death). Let us say that a cohort consists of mostly 75-year-olds. Then it would seem plausible to perhaps consider 25 cycles (where each cycle is one year long).

Once the Markov states, the transition probabilities, and the cycles are all set up you simply run the model. As a result, a distribution of patients across the Markov states for each cycle will emerge. For each of the outcome variables (e.g. costs), one must find a weight associated with each Markov state. If the cycle length is one year, this weight could perhaps be the annual health care cost of a patient in a particular state. The weights together with the patient distribution across states will then produce a value of the outcome variable for that cycle (e.g. the health care costs for the whole cohort). The last step is simply to sum across the cycles. Doing so will create a value of the outcome variable that in some sense represents the impact of the disease course on the variable (e.g. health care costs).

Before doing so, however, consider discounting the values from each cycle. This is simply a way to account for the fact that not all implications of the intervention take place entirely in the present. In short, the farther away from the present a particular implication is

realized, the less of an influence will it have on present decision making. Note also that the higher the discount rate the smaller will the "future" values of the outcome variable become. However, the choice of discount rate is restricted by conventions, typically, in the context of health care interventions, it will lie between three and five percent. However, one should keep in mind that the length of the time horizon will become less significant as the discount rate becomes larger.

Now, in order for the Markov model to inform the comparison of the intervention with the base case (or, perhaps, some competing intervention), it takes an estimate of how the intervention affects the initial distribution (across the Markov states) of the cohort. In addition, it takes data on the distribution holding under the base case. Once these have been acquired, it is simply a matter of running the model twice, once for each distribution, and comparing the difference(s) in the value(s) of the outcome variable(s).

The two alternatives, the base case and the intervention, are often compared in terms of so-called incremental cost-effectiveness ratios (ICER). The idea here is to see how much you would have to pay for an extra QALY (say) if you opt for the intervention rather than the base case.

$$ICER = \frac{Cost_{intervention} - Cost_{base}}{Effect_{intervention} - Effect_{base}} = \frac{\Delta Cost}{\Delta Effect}$$

The numerator holds the difference in the cost variable while the denominator holds the difference in the effect variable. In the present case, the cost variable is a wide cost notion, appropriate from a societal perspective, and the effect variable is the QALY (both these variables will be discussed at greater length in subsequent sections). The ICER so obtained can then be neatly plotted in the so-called cost-effectiveness plane (Drummond et al., 2015).

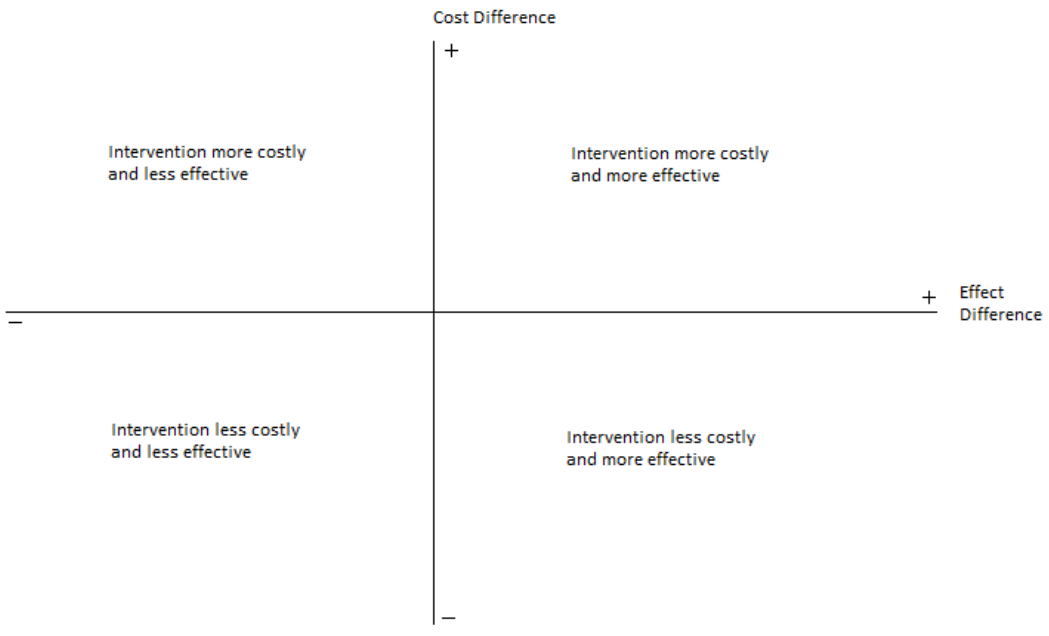

Figure 1 Cost-Effectiveness plane

A version of this plane is given in figure 1. The y-axis charts the difference in costs while the x-axis holds the difference in effect. Consequently, the ICER from the simulation will inhabit one of the four quadrants. For example, a positive cost difference and a positive effect difference will have the ICER end up in the first quadrant. This means that the intervention will have brought about additional QALY albeit with a considerable price tag. With some luck, the ICER will instead end up in the second quadrant. This would suggest that the intervention would bring about more QALY's than the base case and also at a much lower cost. Similar reasoning goes for the third and fourth quadrant in the plane (their respective implications are stated clearly in figure 1).

2.2.2 The modified Rankin scale

The rationale behind the modified Rankin scale (mRS) is the need to describe the variations in the functional recovery of patients suffering from cerebrovascular diseases. In particular, this kind of information is interesting to register at the time the patient is discharged (or transferred) from its index hospitalization. In 1988 it was modified so as to take its present form. There are six different categories of disability called grade 0 through 5. In addition, there is a seventh category designating death. The definition of each category can be found in table 1 (Banks et al., 2007).

The mRS is a disability score that is concerned with so-called basic activities of daily living (BADL) as well as so-called instrumental activities of daily living (IADL). The BADL is exemplified by walking, using the bathroom, dressing, as well as other activities associated with basic self-care (Bonita, 1988). The IADL, in turn, includes managing one's own money, preparing meals, complying with a medical prescription, using public transportation and so forth (de Haan et al., 1995). It is this latter concern with the non-physical aspects of disability that distinguishes the mRS from many other disability measures. It allows the clinician to take into account matters of social and mental functioning. This is noteworthy since such factors have been shown to impact the patient's perception of its own disability (Dennis et al., 2000).

There are many different ways to test the adequacy of a measure such as the mRS. There are studies considering its test-retest reliability, inter-rater reliability, validity and its clinical sensitivity (Banks et al., 2007). With respect to each of these different aspects, the measure seems to perform well. In addition, the measure is widely used as an end point in various RCT's of acute stroke interventions. This speaks for the mRS since it makes it easier to compare how different interventions affect the patients. Proponents like to emphasize the measure's brevity and how simple it is to apply in the context of RCT's (de Haan et al., 1995). However, there are limitations worth considering. It has been documented that non-pertinent factors such as comorbidities (including arthritis and diabetes) (Lai et al., 2002; Lalonde et al., 2004; MacKenzie and Chang,

2002; Nichols-Larsen et al., 2005) and socioeconomic status (Horner et al., 2003) can impact mental and physical functioning. These factors, in turn, will directly impact the mRS score attributed to the patient (Rojas et al., 2005) What is more, stroke incidence is higher among the socioeconomically disadvantaged. It is also the case that comorbidities such as diabetes and arthritis are very common among stroke patients. These last two points are worth stressing since they obviously increase the risk of misapplications of the mRS.

Grade	Interpretation
0	No symptoms at all
1	No significant disability: despite symptoms, able to carry out all usual duties and activities.
2	Slight disability: unable to perform all previous activities but able to look after own affairs without assistance.
3	Moderate disability: requiring some help but able to walk without assistance.
4	Moderately severe disability: unable to walk without assistance and unable to attend to own bodily needs without assistance.
5	Severe disability: bedridden, incontinent and requiring constant nursing care and attention.
6	Death

Table 1 The mRS categories

The categories found in table 1 could be thought of as health states to be represented by Markov states. Recall from the previous section that Markov states have several distinguishing features. First, they should be closely connected with the disease under consideration. This is obviously the case with the mRS categories and stroke since disability is the main health detriment associated with stroke. In addition, conveniently enough, there is the seventh mRS category, i.e. death, that might act as an absorbing state. It is also the case that grades 0 to 5 can be ordered from "better" to "worse". Furthermore, if the mRS categories are thought of as Markov states, the risk of relapsing into a subsequent stroke can serve as the transition probability. It also seems plausible to think of the cohort as degenerating onwards throughout the mRS categories until they eventually reach mRS 6.

Apart from seeming very compatible with Markov modeling, the mRS is convenient in that there has been a lot of academic effort put into relating the categories to various variables that are of interest to policymakers. The three typical variables that figures in evaluations of

health care interventions, for example, are costs, mortality, and quality-adjusted life-years (QALYs). Examples of studies relating mRS to costs are Dawson et al. (2007) and Ghatnekar (2013). Similarly, there are studies that associate each of the mRS categories with a QALY weight. These include Rivero-Arias et al. (2009), Ganesalingam et al. (2015) and Leppert et al. (2015). Lastly, there are examples of research, such as Slot et al. (2008), considering how the mRS relates to mortality.

2.3 Previous research

Unfortunately, it does not seem as if the intervention considered by the county has been carried out elsewhere. Because of this, there are no studies suggesting that it indeed has a clinical effect. However, there are interventions (although not carried out in a Swedish context) that are quite similar. These other stroke care interventions distinguish themselves from the one considered by the county in several different ways. However, there is at least one aspect that appears more prominent, namely, the fact that the intervention considered by the county does not encompass the option of transmitting CT scans from the ambulance to the consulting stroke experts on the receiving end. In spite of this, it will be worthwhile to briefly consider what has been said about these other interventions.

First, there is the intervention associated with the so-called "Brain Attack Team" (BAT) in the U.S (LaMonte et al., 2004). This is the acute stroke response team at the University of Maryland Medical Center which serves as a statewide resource to expedite the treatment of stroke patients. This particular intervention seems to be the most similar to the one considered by the county. It features the possibility to transmit audio and visual information, as well as the patient's vital signs, from the ambulance during transport. The contention is that this would shorten the time to treatment once the patient is known to the emergency medical system. The intervention was successful in that the hospital physician examined the patient en route as well as at the hospital and reached the same conclusion with respect to stroke diagnosis. The authors conclude that the clinical evaluations of the stroke patients were satisfactory but that further improvement in data transfer efficiency and reliability is needed.

On a similar note, Audebert et al. (2008) reports the results from a controlled trial comparing hospital-based stroke consultation with mobile teleconsulting. The former is the conventional way to treat stroke while the latter consists in consultation through a mobile laptop workstation. This intervention does not match the one considered by the county since the mobile unit is not located within an ambulance. However, they evaluated various technical parameters as well as the impact on immediate clinical decisions. There were 127 observations of the

conventional treatment and 96 observations of mobile teleconsultations. The authors conclude that the intervention was technically stable and that it facilitated remote clinical decision making. However, they would like to see improvements of the video quality on the hub side as well as the video transmission to the spoke side.

There is also a general review (Schwamm et al., 2009) of the evidence in support of telemedicine in the context of stroke care. Here, "telemedicine" is defined broadly as "the use of telecommunications technologies to provide medical information and services". However, at the time this review was carried out there was only one instance of ambulance-based telemedicine, namely, the one associated with BAT. The authors concede that the technological preconditions (although outdated by now of course) will support some degree of interactive video and audio communication with prehospital units in transport. However, they hold that the frame rates are unacceptably small and, further, that it is not yet feasible to implement the technology in a larger ambulance fleet.

Seemingly, the first RCT on this subject is reported in Walter et al. (2012). The main endpoint of the trial was time from alarm to therapy decision. The comparison stood between conventional stroke care at a hospital (control group) and prehospital stroke care in a specialized ambulance (treatment group). This ambulance was equipped with point-of-care laboratory, a CT scanner as well as telemedical technology. Note that, since the county's intervention does not include the former two components, one must be careful when translating the results of this RCT to the context at hand. The authors conclude that time from alarm to therapy decision was reduced substantially (median 41 minutes) in the treatment group. Similar results were detected with respect to secondary endpoints such as time of alarm to end of CT, time of alarm to end of laboratory analysis as well as end of intravenous thrombolysis (for eligible patients).

There is also the so-called PHANTOM-S study described in Saifee et al. (2013). This intervention is very similar to the one from the previous paragraph. The main difference lies in that the PHANTOM-S intervention specifies that each ambulance will be staffed by a neurologist, a paramedic and a radiographer. Consequently, this specification takes the intervention even farther away from the one considered by the county. The intervention included 152 subjects and the mean call-to-needle time was approximately 36 minutes shorter compared with conventional emergency treatment. According to the authors, the results suggest that this kind of prehospital treatment is feasible. However, one must note that there has not been any RCT to confirm this contention.

In contrast to the studies considered thus far, Stephanie et al. (2015) describes and comments on a lot of the necessary preparatory work pertaining to telemedical interventions in

stroke care (in the U.S.). The particular intervention they consider has previously been carried out in Germany. There, it managed to demonstrate prehospital treatment of more patients within the first hour of symptom onset. Mobile stroke teams are introduced and equipped with CT scans making the intervention (again) rather different from the county's. According to the authors, the preparatory work will include matters of radiation safety, purchasing supplies, insurance, licensing, staffing, designing protocols to ensure accountability, designing a research protocol and so forth. As a result of the study, it is clear that the major obstacles in paving the way for this kind of intervention are funding, licensure as well as documenting radiation safety protocols.

We also have a very recent RCT carried out in Germany in 2013. The background and results can be found in Ebinger et al. (2014). The author's objective was to consider prehospital thrombolysis for acute ischemic stroke administered in special ambulance units. The contention was that this kind of intervention might have an effect on thrombolysis rates, delay in thrombolytic administration, post-thrombolysis intracerebral hemorrhage as well as 7-day mortality. This study is similar to Weber (2013) in that the ambulance seems to hold similar equipment (CT scanner, point-of-care laboratory and teleradiology system) and staff (neurologist, paramedic and radiologist). In total, 518 patients were included in the primary analysis and the alarm-to-treatment time was approximately 15 minutes shorter in the treatment group compared to the control group. An additional important finding is that the stroke intervention seemed non-inferior to conventional stroke treatment in terms of 7-day mortality and secondary intracerebral hemorrhage.

Lastly, Parker et al. (2015) concerns a non-randomized run-in phase designed to obtain data pertaining to an intervention that would introduce mobile stroke units to the stroke care in the U.S. These mobile stroke units are staffed and equipped just as the ones mentioned in the previous paragraph. The run-in phase stretched over eight weeks and included 24 patients. According to the authors, the main findings of the study are that the agreement between onsite and telemedicine physician, with respect to thrombolysis eligibility, was no less than 90%, but also that important information that can be of help when designing a consecutive and more conclusive RCT has been obtained.

3. QALY's

This third part will hold a thorough discussion of various methodological topics pertaining to the concept of quality adjusted life-years (QALY's). The QALY is an important variable to consider when evaluating a health care reform. To further improve the health of the population at large is perhaps the most natural reason why policymakers would intervene in the health care system. Whether a particular intervention has brought about such an improvement is obviously an empirical matter. However, there is also a wide host of conceptual issues that precedes the empirical investigation. Perhaps none of these issues have an obvious or clear-cut answer but, nonetheless, it is important to highlight them and to make one's standpoint explicit. These matters will be covered in the subsections at hand.

First, the inadequacy of so-called intermediate measures of health will be discussed in **3.1**. In a stroke context, these measures include the time elapsing from onset of symptoms to treatment, and the share of stroke patients directly admitted to a stroke ward. Then, in section **3.2**, a discussion of what it means for a measure of health to be generic and why such measures are important will take place. Lastly, some light will be shed on the matter of how to elicit QALY weights in section **3.3**. This will include a brief introduction to the standard gamble method and the time trade-off method.

3.1 Intermediate measures

As was mentioned a few paragraphs back, in the context of stroke care, intermediate measures of health will be exemplified by the time elapsing from onset of symptoms to treatment, the share of stroke patients directly admitted to a stroke ward, time elapsing from thrombolysis to thrombectomy, or perhaps the share of patients whose diagnosis at discharge is the same as their prehospital diagnosis. It can be seen from the preceding section reviewing previous research that many studies consider these kinds of intermediate measures of health as end points of their investigations.

Now, the contention is that variables such as "door-to-needle time" are somehow related to the subsequent health state of the stroke patient. As suggested by the saying "time is brain", a simple way to salvage as much brain as possible once a stroke patient has been detected might simply be to put the patient under the needle as quickly as possible. Recall from section **2.1** that in case the brain is deprived of a continuous flow of blood it will suffer immediate damage.

This damage will often translate into severe disability and this, in turn, will obviously impact the patient's quality of life. Given that it is the quality of the patient's life that the health care system is concerned with, it would seem reasonable to pay attention to the door-to-needle time brought about by the intervention.

However, it is not always the case that the path from the intermediate measure to the quality of life of the patients is as straight and direct as with door-to-needle time. For example, consider the share of patients whose diagnosis at discharge is the same as their prehospital diagnosis. Perhaps one could argue that when the two diagnoses are the same the patients will have taken less detours within the health care system before being discharged. Such detours could perhaps be detrimental to the patient's health in that they might imply the wrong treatment (postponing the right treatment) and they might also put the patient under a lot of stress. But in the absence of some kind of medical evidence (for example a RCT) that the alleged path from the intermediate measure to patient health in fact exists, the adequacy of the measure will remain tentative.

There is an interesting example of when an epidemiological study has been used to establish the alleged path from an intermediate measure to a more final measure. The medical field is leukemia rather than stroke but the methodological point will be conveyed regardless. Oriana et al. (2013) tried to see whether complete cytogenetic response and major molecular response, both at 12 months, could predict overall survival after treatment of chronic myelogenous leukemia. Here, the former two variables are intermediate measures while the latter, survival rate, is the final measure. The authors undertook a systematic review of existing observational studies and managed to convince policymakers of the alleged association between intermediate and final measure.

In some cases, of course, it will be too expensive or too tedious to conduct a RCT or a systematic review of existing observational studies. Then, one could resort to lesser forms of evidence such as expert opinions or common sense judgements from the profession. However, in doing so one must be aware that lesser forms of evidence will translate into lesser credibility of the economic evaluation.

A second concern is that intermediate measures, such as door-to-needle time in the case of stroke, are often non-generic. The importance of generic measures will be covered more in depth in the next section. However, in short the main problem with non-generic measures is that they often render interventions incomparable. This is because such measures are often disease-specific. To see this, consider some stroke care intervention the evaluation of which includes door-to-needle time as an intermediate measure. Consider also the leukemia example from the

previous paragraph where so-called "major molecular response" served as an intermediate measure. These two interventions might very well be competing in the sense that policymakers must choose to implement one and discard (or at least postpone) the other. Unfortunately, policymakers will have difficulties making an informed decision since they cannot easily translate one of the measures in terms of the other. Put differently, there is no common dimension (apart from costs of course) along which both interventions can be compared.

3.2 Generic measures

When carrying out an evaluation of a health care intervention there are a wide variety of measures to consider. In the previous section, there was an implicit distinction made between intermediate measures and final measures. With the last section serving as a background, it will make sense to think of intermediate measures as being non-final albeit having an alleged connection to a final measure. A final measure, in contrast, should be construed simply as a measure that is directly concerned with the patient's health. Put differently, it explicitly displays or depicts the health state of the patient.

This distinction, now made explicit, can be neatly exemplified by some of the measures from the past section. Think again of the so-called "major molecular response" from Oriana et al. (2013). This measure is obviously not final since it does not explicitly tell us something about the health state of the leukemia patient in question. However, it can be thought of as intermediate since it has an alleged connection to a final measure, namely, survival rate, or perhaps QALY's. The latter two measures are final in that they tell us something about the health state of the patient in a direct and explicit manner. In a similar fashion, time from onset of symptoms to treatment is to be construed as intermediate since it is obviously non-final although in a stroke care context it could be related to QALY (a final measure).

In addition to this distinction, it will be worthwhile to also distinguish between generic and disease-specific measures. A generic measure is a measure that can be deployed in any intervention regardless of the medical field. Put differently, they do not focus on the impacts of a particular disease. Rather, a broad range of dimensions of quality of life will be taken into account. These various dimensions are exemplified by physical functioning, social functioning, pain, and mental well-being. The most widely used generic measure is perhaps the QALY measure (originating from Klarman et al. (1968)). Another generic measure is the disability adjusted life year (DALY) from Murray and Lopez (1996).

Disease-specific measures of health, in contrast, only accommodate the main quality of life impacts of a specific disease. This means that they will be narrower and consequently hold

less information compared to a generic measure of health. It is also the case that they cannot be applied outside of their original medical field. Two prominent examples of disease-specific measures are the EORTC instrument (Fayers and Bottomly, 2002), applied in the field of cancer, and the St George's Asthma Quality of Life Scales (Jones et al., 1992) which can be used to assess the impacts of asthma.

The concepts of intermediate measure and final measure are mutually exclusive. The same is true of the second pair of concepts, namely, disease-specific measure and generic measure. However, it is important to note that a final measure could be either disease-specific or generic. Similarly, intermediate measures can also be either disease-specific or generic. As an illustration of this, consider the intermediate measure from Oriana et al. (2013) mentioned earlier, namely, major molecular response. This measure is disease-specific since it cannot recur in every medical field. The same is true of the intermediate measure time from thrombolysis to thrombectomy since it will not recur outside of a stroke care context. In contrast, time from onset of symptoms to treatment, an intermediate measure for sure, is generic in that it obviously can be applied in any medical context.

In a similar fashion, there are final measures that are disease-specific. Examples include the measures from a paragraph back, namely, the EORTC instrument and the St George Asthma Quality of Life Scale. The former, as you recall, is tied to the cancer field while the latter is associated with asthma. As for final measures that are at the same time generic, one could think of the QALY measure or perhaps a measure of mortality. These are final in that they concern the health state of the patient in a direct manner and they are generic in that any major disease will obviously have an impact on the QALY and the mortality rate of the patients.

The taxonomy that has just been drawn out will hopefully be worthwhile considering in situations where the policymaker must decide on what variables to register and what variables that should serve as end points of economic evaluations. In order to make the relations between the concepts more transparent and tractable, the above examples have been summarized briefly in the following table.

Table 2 Measure taxonomy

	Final	**Intermediate**
Disease-specific	EORTC	Time from thrombolysis to thrombectomy
Generic	QALY	Time from onset of symptoms to treatment

Now, when choosing the end point(s) of an economic evaluation, one should go for a measure that is (i) final rather than intermediate, and (ii) generic rather than disease-specific.

The reason for (i) was stated in the last section. In short, the intermediate measure itself will not suffice as an indicator of the health of the patient. At best, it could serve as a mediator between the intervention and some proper measure of patient health. However, in order for policymakers to have faith in such a nexus there will have to be proper medical evidence, perhaps a RCT, in support of the contention in question.

As for (ii), the main reason is that policymakers will want to be able to compare interventions even though the interventions concern different medical fields. However, this matter of comparability has two different aspects. The first turns on the accountability of the policymaker. It is important for the sake of consistency of the decisions made that it is possible to compare across different points in time and across the various patient groups that are affected by the decisions. Secondly, some interventions will have side-effects that can only be detected in other medical fields. For example, interventions in the asthma care can be partly evaluated by means of the St George Asthma Quality of Life Scale. However, in case the intervention causes negative side-effects outside of this medical field, something in the field of cardio vascular diseases say, they will not be detected since they would be out of reach of the asthma scale. In contrast, they would not go undetected if instead the evaluator resorts to a generic measure like the QALY.

3.3 The elicitation of weights

Since the QALY seems an appropriate end point of an economic evaluation in that it is both final and generic, it will be worthwhile considering the particulars of incorporating it into a Markov model. In section **2.2.1** where the Markov model was first introduced, it was said that weights must be constructed for each outcome variable. These weights are then to be attached to the Markov states, one weight per state, and then you go on to simulate a value of the outcome variable in question.

However, there are different ways of eliciting these weights. It is important to highlight this fact since the choice of elicitation method can affect the size of the weights (Suarez-Almazor & Conner-Spady, 2001) and, consequently, the outcome of the evaluation.

Before considering methods of elicitation, note that it will also matter from whom the weights are elicited. A common distinction is that between hypothetical and experience-based health state valuations (Burström et al., 2013). Note that, in the present terminology, the health state valuations could be seen as the QALY weights associated with the Markov states.

Experience-based QALY weights are elicited from individuals that have actually contracted the disease or condition in question and are afflicted by it at the time of the elicitation. Hypothetical QALY weights, in contrast, are elicited from a representative sample of the general population that is not afflicted by the disease at the time of the elicitation.

The idea behind experience-based weights is that the perceptions of actual patients are probably the best proxy of how people in general would feel should they contract the condition. It might be difficult for healthy individuals to anticipate what the condition would be like. However, some conditions will have afflicted the patient for a long time making it difficult for the patient to recall what it was like to be healthy. If there is no recent experience of full health, the patients might have adapted to their condition, thinking of it as the "normal" health state rather than a degenerate one, and consequently assign it a weight that is too large (Brazier et al., 2005; Dolan, 2008; Dolan and Kahneman, 2008; Dolan, 2011)

One real-life example is where colostomy patients have been asked about their condition. They respond that they are content with their colostomy and also that they expect to remain content if the colostomy were to be removed. However, once the colostomy in fact is removed they recall their previous health state as horrible and state that they would be willing to pay a lot of money to avoid it (Smith et al., 2006).

On the other hand, hypothetical weights could perhaps be more appropriate since we are all tax payers and potential patients and therefore all of us should have a say in prospective health care interventions (Gold, 1996). An important difference here is that there will be no "adaptation effect" skewing the elicitation. However, in order to talk about an elicitation being skewed, in one direction or the other, one must first have some kind of standard in mind. Unfortunately, one cannot think of the hypothetical weights as being the standard without begging the question. Given such a standard, however, perhaps the hypothetical weights, rather than touching this standard, will be skewed below due to exaggerated fears of losing one's health. As it turns out, empirical studies suggest that the experience-based weights are indeed somewhat higher than the hypothetical weights (Mann et al., 2009; de Wit et al., 2000; Ubel et al., 2003). This is important to keep in mind when the evaluation of the county's intervention is to be carried out.

Having decided on if the weights should be hypothetical or experience-based, there are three methods of elicitation that are the most widely used. Again, there is no clear-cut answer to which of these is the most appropriate. But since there are many studies, e.g. Bass et al. (1994), Dolan et al. (1996), showing that the choice of elicitation method will affect the weights, it is important to at least be aware that alternatives exist and also to make one's choice explicit.

First, there is the rating scale approach and its variations. This method is often perceived has the most basic and it can be thought of as having two steps. The first consists in asking the subject to order a set of health states from most preferred to least preferred. Secondly, the subject is asked to place the health states on a scale and be mindful of the spacing they put in between. Ideally, the spacing should correspond to the subject's perception of how preferable one health state is compared to another. Put differently, if the subject perceive health state A as being twice as preferable as health state B, and three times as preferable as health state C, then A should be twice as high on the scale as B and three times as high as C meanwhile B should be placed slightly above C.

The variations of this method are called the rating scale, category scaling and the visual analogue scale (VAS). It is called a rating scale when the subject is presented with a scale filled with integers (often) from 0 to 100. Category scaling, instead, is when these integers are replaced by a set of categories. There are often 10 or 11 categories and the subject is encouraged to think of these as being equally spaced. Lastly, the VAS is simply a line on a page that is 10 cm in length. It must also be that the end points of this line are clearly marked out.

Another way of eliciting weights is through the so-called standard gamble method. Here, it is important to distinguish between those health states that are chronic and those that are temporary. When considering a chronic health state, the subject is presented with two alternatives. The first is a treatment with two outcomes. Either the subject will, with probability p, regain full health and live for another t years. Alternatively, with probability $1 - p$, the patient will die immediately. If the patient does not choose treatment, it will instead remain in the chronic state for t years. The idea is then to vary p until the subject is indifferent between treatment and no treatment. Once this particular p has been found, it is to be considered the QALY weight for the health state in question.

Lastly, there is the method of eliciting weights called the time trade-off method (TTO). This method comes in many variations but the most basic version draws a sharp distinction between chronic and temporary health states. In case of a chronic condition, the patient is asked to choose between two alternatives. The first is to remain in the chronic health state for the remainder of its (expected) life time t (i.e. until death) while the second is to be healthy for time $x < t$ followed by death. The time x is then to be varied until the patient is indifferent between the chronic state and being healthy. The weight for chronic state is then given by $h = x/t$.

As for temporary states, the patient is asked to value these in relation to each other. Either the patient remains in state i for time t (and then reverts to the healthy state), else the

patient will be in state j for time $x < t$ followed by the healthy state. Again, x is varied until the patient is indifferent at which $h_i = 1 - (1 - h_j)x/t$.

4. Mortality

This chapter will elaborate on another of the key components of the Markov simulation, namely, the probability of transitioning to the state of death. Recall that, in terms of the mRS, the state of death corresponds to the sixth mRS category and that, and the end of each cycle, there is a chance that the patient will transition to that particular category. Recall also that, as the simulation unfolds, this particular transition probability will be what determines the number of life years accumulated by the simulation. These, in turn, will be necessary for the calculations of both QALY's and costs.

First, there will be a section devoted to how the mRS is related to mortality. Secondly, there will be a section on how to incorporate mortality into a Markov simulation.

4.1 mRS and mortality

In order to make the Markov simulation as realistic as possible, one will need transition probabilities, from each of the mRS categories to the sixth one, that are somehow anchored in reality. One way to accommodate this is to first consider so-called life tables for the patient population in question. There are many different kinds of life tables but typically they will be so-called periodic life tables or else they will be so-called cohort life tables. Periodic life tables will contain the probability of death across different ages for the current year. The cohort life tables, in contrast, will hold the probability of death, associated with people from a particular cohort, over the course of their whole life time. In case of the latter, one simply chooses a cohort that fits the stroke population under consideration (for example men or women at the age of 65). These kinds of numbers can easily be obtained from e.g. Statistics Sweden.

Then, once an appropriate life table has been obtained, the probabilities should be altered by means of so-called risk ratios associated with each of the mRS categories. Basically, a risk ratio for a given mRS category will say how much greater the risk of dying is, if one inhabits that particular category, compared to the general population. For example, say that a patient is 75 years old and that currently the patient's level of disability is classified as mRS four. Say also that the risk ratio associated with mRS four is 2,5 and that the underlying mortality rate of a 75-year-

old is 0,3. In order to find the risk of dying associated with the patient in question, one simply multiplies the underlying mortality rate of 0,3 with the risk ratio of 2,5 and obtains 0,75. This means that the patient in question suffers a risk of dying (within the next year say) equal to 75%.

As mentioned above, the underlying mortality can be obtained easily from Statistics Sweden. However, the risk ratios cannot. Luckily, there are studies, such as Slot et al. (2008), that have attempted to extract these ratios as a part of a broader stroke related study. The aim of this particular study was to see if patients' degree of disability (so-called functional status) had an impact on their long-term survival rate. More specifically, it was the degree of disability six months after an ischemic stroke that was presumed to have an impact. The study design was a prospective cohort study with three different cohorts. In total, there were 7,710 patients participating in the study and they were all registered between 1981 and 2000.

The authors found that the level of disability indeed had an effect on long-term survival. This effect grew stronger with the level of severity of disability. As an example, across two of the cohorts the median survival for patients ending up in mRS zero or one was approximately 13 years. In contrast, patients in the most severe category, mRS six, had a median survival of no more than 2,5 years. As for risk ratios, they vary from as low as 1,2 for the least severe disability levels, to as high as 2,78 for the most severe.

Admittedly, this might seem like a rather complicated procedure. Furthermore, perhaps the transition probabilities so obtained are not sufficiently realistic. The reason for this is that eliciting risk ratios from some patient population, other than the one that the county is responsible for, might skew the resulting transition probability. For example, the study mentioned above concerns patient cohorts from the U.K. Although the health care institutions in the U.K. are not all too different from the ones in Sweden one cannot exclude the possibility of systematic differences in patient characteristics.

One way to avoid this risk of skewed transition probabilities would be to instead elicit them directly from a Swedish patient cohort. This way the outcome variables at the end of the simulation would become more realistic. One way to facilitate the acquisition of these probabilities is to try and exploit the database called Riksstroke. Although not presently available, perhaps their administration can be convinced that these kinds of probabilities are important for the evaluation of stroke care interventions. If so they could be routinely elicited from the Swedish population and because of this the evaluation of various stroke care interventions could be made much more rigorous[3].

[3] For details on how to properly extract transition probabilities from cohort studies see Drummond et al. (2015), chapter 10, and also Briggs et al. (2006).

4.2 Mortality in the Markov model

In general, one would think of the mortality rate of some individual as something that would increase as the individual grows older. Put differently, it would be strange for an individual to have the same mortality rate at the age of 15 and 65. However, this is exactly what is presumed in the Markov model, at least the version that was sketched in section **2.2.1**. In terms of the model, this means that the transition probabilities going from the various mRS states to the death state do not change as the simulation progresses. Any given cycle will be associated with exactly the same mortality rate as the cycle previous (or subsequent) to it.

As a consequence, one could expect the simulation to exaggerate the actual number of life years that would be brought about by the intervention. Had the mortality rate been allowed to increase as the cycles unfolds, the argument goes, a greater share of the cohort would have died before the end of the simulation. This, in turn, would reduce the amount of QALY's that the intervention (presumably) would result in and, everything else being equal, the case for the intervention would be undermined. Something similar, of course, happens to the cost variable. Perhaps it will be inflated due to the fact that a larger share of the cohort will remain alive albeit in a poor state of health which demands a lot of health care resources. At the same time, one could expect costs to be reduced since the intervention will also result in more patients remaining in the healthier parts of the mRS scale (meaning less demand of health care services).

Although there are ways to allow for transition probabilities (such as the mortality rate) to change across cycles[4], they seem rather unnecessary. The reason for this is that the Markov simulation will always compare the intervention with a base case (or "status quo"). After having run the simulation as described in section **2.2.1**, one will have a cost and QALY variable for the base case as well as for the intervention. One will then proceed to form two ratios, one for the base case and one for the intervention, which will tell how the two regimes differ with respect to effectiveness. This means that whatever bias there is in the number of life years accumulated by the simulation will be afflicting both scenarios equally. Put differently, the lack of time sensitive transition probabilities will be present both in the simulation of the base case as well as the simulation for the intervention. The alleged bias will therefore cancel out since the comparison will be in terms of how much more (or less) effective the intervention is compared to the base case.

[4] Briggs et al. (2006).

5. Costs

This chapter holds discussions of various topics that somehow pertain to the cost variable that will be brought about by the intervention. The discussion will be made slightly more tractable if a distinction is drawn between direct and indirect costs. This distinction is simply for the sake of readability and comprehension. Each cost category will encompass yet other kinds of costs and each of these will be clearly delineated and commented on in what follows. Hopefully, the considerations presented in this chapter will be informative when trying to produce a cost weight associated with each of Markov states in the simulation.

Note that the notion of costs used in the following sections is wide to say the least. As was mentioned at the outset, the vantage point of this essay is the societal perspective. As a consequence, certain components of the cost notion might seem alien from the county's perspective (which is presumed to be somewhat narrower). However, part of the rationale for this wider perspective is that it is much easier to disregard things deemed unnecessary than it is to include something (perhaps important) one has never heard or thought of before.

Lastly, since the costs (however defined) of the intervention are not properly charted at the time of writing, the reader must not expect this section to hold any detailed information on the cost structure. Instead, the subsequent sections will be more methodological in nature. Put differently, they will sketch the conceptual groundwork that precedes any description of the actual costs.

5.1 Direct costs

Now, direct costs will here be construed coarsely as the resources depleted by the health care sector as a consequence of the intervention under study. Here, one will find three subcategories, namely, variable costs, fixed costs, and overhead costs. These types of costs should be contrasted with the indirect costs. Typically, it is a concern with this latter kind of cost that distinguishes a societal perspective from a regional (or county) perspective. Consequently, in order to adopt a narrower than default scope, one way is to simply disregard the indirect costs and focus solely on the direct ones. The direct costs are typically the ones that are the easiest to apprehend. It will often come down to making a list of the ingredients necessary to carry out the intervention. All the resources devoted to organizing and operating the intervention must be included in this expense item.

5.1.1 Variable costs

The variable costs of the intervention are the costs that somehow depend on the number of patients treated under the new regime. This will include the wage of the personnel operating the new equipment in the ambulances as well as the wages of the stroke experts consulting at the receiving end of the telecommunication line. Perhaps one could object that the ambulance personnel are already hired since there must be ambulance personnel also in the base case scenario. The same could be true of the stroke experts consulting the personnel in the ambulance. Why then, the argument goes, should these expense units be considered at all? Once the two Markov simulations are done, the argument continues, the base case and the intervention will be compared and the aforementioned costs will cancel out.

This will indeed hold true as long as the ambulance personnel has not been given a raise (translating into increased costs for the employer), perhaps due to the extensive training necessary to properly operate the new equipment, or as long as the employed is not forced to neglect other tasks in order to operate the new ambulance properly (forcing the employer to perhaps hire additional personnel to see to these neglected tasks).

Another important variable cost is any additional labor responsible for maintenance of the new ambulance equipment. Just as with the example from the previous paragraph, it is important to note that it is the additional labor that is interesting. Presumably, the ambulances already demand a certain amount of maintenance. This maintenance cost will be covered in both the base case scenario and the intervention scenario and, therefore, it will cancel out at the end of the simulation. In addition to these personnel costs, one must also consider the cost of various medical supplies that are depleted with the treatment of a patient. However, if it turns out that the intervention consumes the same worth of medical supplies as the base case, this expenditure item can (again) be neglected.

In general, for the sake of an economic evaluation (by means of a Markov model), what is relevant to make note of when it comes to variable costs is the additional variable costs. Any other costs, the ones that occur in both cost columns as it were, will cancel out once the simulations are done and the comparisons are made. Therefore, the analyst will save a lot of effort by neglecting this latter kind of costs regardless of how important they might seem.

5.1.2 Fixed costs

In addition to the variable costs there are costs that do not depend on the number of patients treated. These costs are called fixed costs and are exemplified by all the new equipment that comes with the intervention, but also by the costs associated with training the personnel to

operate this equipment. The distinguishing feature is that these fixed costs occur only at one point in time while the variable costs recur (annually say). They represent the major outlays made, often in the beginning of the implementation of the intervention, in order to get the key capital assets in place. The assessment of the variable costs from the last paragraph seemed rather straightforward. In contrast, there are two important methodological notes to make when it comes to fixed costs.

Firstly, resources spent on the capital assets of the intervention will be tied up in those assets for as long as they are not sold. Put differently, the opportunity cost of the investment will be stuck in the asset in question. This means that the policymaker will not be able to invest the money in some other venture that could have brought positive benefits. Say that the policymaker invests in a piece of land. Having done this, the opportunity to invest that particular amount of money will be lost and will not return until the land is sold. Typically, the matter of lost opportunities to invest is accommodated by applying an interest rate to the amount of resources invested[5]. This interest rate is to be equal to whatever discount rate is used (see section **2.2** for details on discounting).

Secondly, there is the matter of the depreciation of the asset. Any asset invested in will depreciate over time meaning that it will somehow be worn down which will lead to its monetary value diminishing. In the case of a car for example, the engine and the tires and all the other hardware will wear out over time. Eventually, what is left of the car will be so worthless it will end up at the scrapheap. Similarly, one's wage will depreciate as time passes due to a steady positive rate of inflation. Unless there is an appropriate raise so as to counteract this devaluation, the wage will eventually be worth very little in real terms.

There are many ways to measure or value these fixed costs. However, the one recommended by Drummond (2015) is to "annuitize" the capital outlay of the concerned asset. This amounts to figuring out the annual cost of the asset over its useful life. If one opts for this method, one will automatically incorporate the opportunity cost as well as the depreciation cost (previously discussed) into the annual cost. Details on how to best carry this method can be found in chapter seven of Drummond (2015).

5.1.3 Overhead costs

The last of the direct costs is the so-called overhead costs. As will become evident, these are not entirely separate from the previous two categories. Rather, some overhead costs are variable costs while some others are fixed costs. In general, overhead costs are the depleted resources that at

[5] See chapter seven of Drummond et al. (2015) for further details.

the same time benefit many different departments or programmes within the health care sector. Typical examples include the costs for central laundry, general hospital administration, power, medical records and cleaning of hospital facilities. It is important to try and discern how much of overhead costs are exhausted by the intervention and the base case respectively. The reason for this is that the two programmes might deplete overhead resources at different rates. If, say, the intervention will deplete less than the base case, then, all else held constant, the intervention is obviously to be preferred.

Although there does not seem to be any consensus on how to best estimate the overhead costs associated with a given intervention, the most common way to do so is to use so-called marginal analysis (Drummond, 2015). The way this is to be done in the present case is to, first, chart the overhead costs associated with the base case. Then, see if the introduction of the intervention would affect these overhead costs. Given that there indeed is an influence, one would of course also make note of whether the particular overhead cost that is influenced will increase or decrease. Put differently, try to discern how the addition of the intervention would affect the overall activity of the hospital(s) in question.

5.2 Indirect costs

The subsequent discussion will revolve around costs that are typically less straightforward to deal with. In spite of being difficult to deal with, these cost items are no doubt important to take into consideration. Because there does not seem to be any clear-cut way to deal with these important factors, they should perhaps be seen as unavoidable caveats that serve to deflate the credibility of evaluations of health care interventions. The first such factor is called unrelated costs while the second is called productivity costs. Both are especially interesting since the scope suggested in the present context is the wide societal perspective rather than the narrower county (or regional) one.

5.2.1 Unrelated costs

The first expense item to be considered in this section is the co-called unrelated costs. This term is to be read in its most literal sense. The unrelated costs of an intervention are the health care costs that, at least at first glance, might not be easily seen as created by the intervention in question. One example is the future health care costs that come from extending the life of a patient. Undoubtedly, these additional life years must be seen as a result of the intervention. This is at least the case when one considers its benefits. Put differently, the fact that the patient lives longer is usually put on the pro side when trying to decide whether the intervention is worthwhile or not.

However, additional life years could mean many different things in terms of future health care costs. Let us consider stroke as an example. Surviving a stroke at the age of 65, because of the improved pre-hospital care suggested by the county's intervention say, would most likely lead to disability. However, the patient would presumably suffer from a more severe degree of disability had the intervention not been implemented. As per usual, the lack of perfect health will entail a need for health care services. Now, is this presumed future need of disability related health care services to be considered a direct consequence of the intervention? Well, if the additional QALY's are deemed related then perhaps so should the cost of additional health care services.

But what if the stroke patient whose life has been prolonged contracts cancer? This condition has little to do with stroke but will nonetheless demand very costly health care services. At the same time, it is hard to deny that had the patient not survived the stroke at the age of 65 these subsequent cancer related problems would never have arisen. The problem of clearly delineating which health care costs are related and which are unrelated is especially pressing in the case of stroke. This is because stroke patients are often at least 65 years of age. Many conditions, such as cancer or diabetes, have a much higher incidence among the elderly. As a consequence, the distinction between related and unrelated health care costs, in spite of being blurred and murky, will have to be appealed to more frequently throughout the evaluation process. Each of the (age-related) expenditure items will have to be deemed either related or unrelated. The fact that such judgement are difficult to make suggests that different economists will have different opinions. This fact, in turn, is important to keep in mind since it hampers the comparability of evaluations of different stroke care interventions.

This problem of distinguishing related from unrelated costs is widely recognized and one can find discussions in for example Gold (1996) and Schulman (1991). An especially strong standpoint can be found in Weinstein and Manning (1997). There, it is suggested that the practice of weeding out unrelated costs from evaluations will be hopelessly inconsistent unless clear-cut definitions of "related" and "unrelated" are agreed upon beforehand by practitioners.

Now, until such definitions have been laid down, practitioners will have to settle for a second-best solution. Often this amounts to considering a variable such as the annual per capita health care expenditure. One could also make this variable sensitive to the age of the patient cohort in order to make the simulation more realistic (Drummond, 2015). This way of dealing with the matter will leave the conceptual problem of distinguishing related from unrelated costs unresolved. Note also that it will circumvent the matter of whether unrelated costs, however defined, should be taken into consideration.

Lastly, it is interesting to note that (alleged) unrelated costs associated with prolonging the life of the patient cohort might not always have much of an impact on the outcome of an economic evaluation. This is partly because of the application of discount rates in simulations. Recall that the farther away in time a particular cost is realized the heavier will it be discounted. One example where the effect has been small is given in Drummond et al. (1993). Adding an expenditure figure for health care costs to be expected in the additional life years only changed the cost per life-year gained by 2%. In contrast, Daly et al. (1992) discovered that a similar change in their cost analysis increased the cost per life-year gained by as much as 10%. This contrast serves to show that the matter of unrelated costs can be of great importance for the outcome of the evaluation in spite of discounting.

5.2.2 Productivity costs

Whenever an individual contracts a disease or undergoes some kind of health enhancing medical therapy, it is often the case that the individual's labor force participation will be affected somehow. In fact, in some cases, this effect might extend to the patient's family or friends. For example, after having undergone some medical therapy, the patient will perhaps be unable to carry out certain chores and will need help from relatives or friends. Perhaps the chores will be so demanding that one cannot remain at work for as long. This means that not only will the patient have to abstain from working (in order to recover properly say) but the relatives or friends helping out with the chores will also have to abstain from working to some extent. Admittedly, the effect might as well go in the opposite direction: perhaps the therapy will alleviate the patient from its condition, this will enable the patient (and whatever relatives or friends that have been helping out) to devote more time to wage labor.

There is a similar concern with life-saving or life-extending therapies. If an individual survives a stroke because of improved pre-hospital stroke care the individual will perhaps be able to work longer. The same will be true even if the patient's life is not at stake. Improved pre-hospital stroke care could lead to less severe disability meaning that the patients will perhaps not be as unproductive. Note also that it is not only the private income of the affected individuals that is a concern here. The private income of the individuals will generate tax income as well so, consequently, if the former is somehow reduced or enhanced so is the latter.

However, in spite of seeming like an important aspect of health care interventions, there is widespread disagreement among economist on how to deal with these productivity effects. In order to clearly see why this is so there is an example given in Drummond (2015) that is worth considering. Suppose the medical field is mental health and that there is an intervention that must

be compared to a base case. The latter requires institutionalization of the patient for a certain period of time. The former, in contrast, will amount to the patient staying at home more since it consists in the deployment of community-based psychiatric nurses in conjunction with outpatient hospital visits. By assumption, the two scenarios' effect on the patients' mental health is the same.

Now, in terms of actual costs to the health care system, the intervention is assumed to be the more expensive. However, the number of work days lost is considerably lower with the intervention since the patient cohort will not be institutionalized. This will obviously result in lower productivity costs for the intervention (compared to the base case). The question asked by the economist carrying out the evaluation is now whether these gains in productivity should be deducted from the costs associated with implementing and maintaining the intervention. At first glance, it would seem as if the economist indeed must include these productivity effects into the analysis (especially when there is a commitment to the societal perspective). However, there are a number of considerations that has to be made before this turn is taken.

First of all, one might think that the base case means a loss in productivity since individuals from the patient cohort must abstain from work. However, it might very well be that the resulting vacancies will be filled by other (unemployed, perhaps) members of the society. This would then counteract the presumed loss in productivity and leave the intervention and the base case on equal footing.

Secondly, there is the productivity effects associated with the opportunity costs of the intervention. The intervention will imply costs on the health care system that most likely will displace other interventions, perhaps in other medical fields, considered by the policymakers. Supposedly, this will have a negative health impact on the concerned patients in these other medical fields. These opportunity costs, in turn, will of course contain productivity effects by the same argument as the more direct costs discussed above. These productivity effects, occurring in other medical fields, then, can obviously only be disregarded at the expense of the policymaker being consistent.

Lastly, as for the estimation of these productivity effects, there does not seem to be a consensus. Typically, they are estimated by gross earnings where benefits as well overhead costs for the employer are included. This method is sometimes referred to as the human capital approach. In addition, there are a number of ways to try and include the productivity effects pertaining to those not currently in aid employment (for example homemakers). These include considering the cost of replacing the individual absent from the work place, the average wage of the individual in question, and the opportunity cost of the production that would have been realized were the individual not absent.

The lack of consensus stems from the fact that each of these methods could be thought of as overestimating the actual productivity effect. From a societal perspective, if a worker is absent for a short period of time the lost production can be compensated either in the meantime by a co-worker or by the same worker once the worker returns to the workplace. Another reason is that, since all jobs include tasks that are more or less important, and since it is often the less important ones that have to be neglected in case of absence, the average wage will overestimate the productivity loss at the margin. These concerns are akin to the more general thought of the transaction costs and other inefficiencies that often comes with any allocation of resources. For example, if one were to close a large mental institution there would be many instances during the closure process where the wards would be under occupied since such a large closure cannot take place overnight.

As a last note, it is important to stress that the policymaker cannot take the productivity costs of an intervention lightly. The most important reason for this is that productivity costs are so unpredictable. Put differently, productivity costs could go in either direction. This was illustrated by the examples from the preceding paragraphs. From the policymaker's perspective, it would have been much simpler if their inclusion either made the intervention's prospects (with respect to cost efficiency) only better or only worse. Let us say that the inclusion of productivity costs in the analysis always made the intervention seem more expensive. One could then choose to neglect the costs, save a lot of time and effort otherwise devoted to their estimation, and simply keep in mind that the intervention's cost estimate is slightly inflated and perhaps make the suitable adjustments at a later stage.

6. Concluding Remarks

In what follows, the main points of the preceding chapters will be briefly summarized. Recall that the purpose of the present study is to suggest and sketch how a particular stroke care intervention, considered by the county of Scania, could be evaluated. In addition, a number of methodological issues, pertaining to the various variables figuring in such an evaluation, are to be properly drawn out.

First, there was an introduction to a means of evaluating the reform aforementioned. More specifically, the practice of Markov modeling was suggested with the county's purposes in mind. Perhaps the strongest argument in favor of choosing a Markov simulation is the frequent use of the so-called mRS measure of disability. This apparatus can easily be thought of as Markov states in the simulation. Another argument is that the long time horizon often associated with stroke is very easily simulated in Markov modeling. The long time horizon is called for in order to properly track the impact of stroke. For example, it is common to relapse into consecutive strokes after having suffered a first and the most widespread complication is disability which often is a life-long state of health.

The application of the mRS is wide in the sense that there is a lot of academic effort already put into finding appropriate weights for potential outcome variables (such as costs and QALY's). Additionally, the nation-wide database Riksstroke will soon start to properly register data on how discharged patients are distributed across the mRS spectrum[6]. The fact that the mRS categories are so widely used will make the evaluation more comparable to evaluations made in other (yet similar) contexts. This not only makes the evaluation more scientifically interesting but also makes it easier for the tax payers to hold the county accountable for whatever decision is made. Increased comparability will also benefit the county in that whatever decision is made can potentially be made with more confidence since there are many more studies that can serve as reference points.

In the same chapter, there was an overview of similar interventions (and their evaluations) carried out in different contexts. Although these were all clearly within the telestroke field, they distinguished themselves from the intervention considered by the county. What stood out the most was the fact that the more similar interventions all had some kind of mobile CT scanner installed in the ambulances. This difference is not to be neglected since that piece of equipment will make a big difference when it comes to relaying information to the consulting

[6] Private correspondence with Magnus Esbjörnsson (Hässleholms Sjukhus).

stroke experts. Because of this, albeit these other interventions seemed to show great promise, there is no study lending empirical support to the contention that the county's intervention will indeed have a clinical effect.

This fact is of course detrimental to the credibility of the policymaker although it is not completely devastating. One could still carry out the Markov simulation although one would have to make a range of assumptions as to how the presumed effect would turn out. Then, if the intervention turns out to be cost-effective regardless of the size or direction of the hypothesized effect, one could perhaps conclude that the intervention indeed would be worthwhile.

However, because of the Markov states being mRS categories, the best way to construe the presumed effect would be in terms of distributions across the mRS. Put differently, if the county decides to oversee (say) an RCT that hopes to discern a positive clinical effect of the intervention, mRS distributions would seem a very suitable outcome measure. Such a study would make the intervention easier to simulate but also make the modeling results more credible.

The subsequent chapter considered various methodological issues pertaining to the QALY variable and also its relation to the mRS. The main part concerned the distinction between intermediate and final measures as well as that between disease-specific and generic measures. There was an argument suggesting that one should always try to opt for measures that are final and generic. This would then serve as the main motivation for embracing the QALY as one of the outcome variables to be used in the simulation.

The next section dealt with the matter of eliciting QALY weights. Apart from briefly describing the various methods currently available for doing so, there was some space devoted to drawing out the problems pertaining to the distinction between hypothetical and quality-based QALY's. These issues have no clear-cut answer but since the choice of method have been seen to impact subsequent results, it is important that the policymaker is aware of them. It was also indicated that there in fact exists several studies that have already extracted QALY weights for the mRS categories. However, one must note that these are not from a Swedish health care context and that this obviously is less than ideal.

Following this was a short chapter on mortality. This is an important component of any Markov simulation. It takes the shape of transition probabilities going from each of the Markov states to the so-called absorbing state. In short, different means of coming up with empirical counterparts to these transition probabilities (mention was made of life tables as well as risk ratios) were introduced. Unfortunately, the study which had produced the risk ratios for mRS categories had not made use of a Swedish patient cohort. Again, this is less than ideal. However, perhaps it could serve as to suggest what kind of study the county should administer in order to

get better data for their evaluation.

Lastly, there was a chapter devoted to the cost variable. Firstly, there was a distinction drawn between direct and indirect costs and the chapter proceeded with the details of the former kind of costs. For the most part, this kind of expense item is unproblematic. The exception is of course the overhead costs which can be quite complicated to deal with. In addition, the general (methodological) point that when collecting data on the costs associated with the intervention and the base case respectively, one should disregard any expense item that, as it were, occurs in both columns. In other words, if one and the same cost item occurs in both scenarios, say whatever medical supplies are typically depleted in pre-hospital stroke care, one can disregard this particular cost completely. The reason for this is that once the two simulations for each scenario are completed the economist will subtract the one from the other. What will make a difference, then, are the additional costs that the more expensive scenario implies in contrast with the cheaper one.

As a last note, the more complicated cost items, namely, unrelated costs and productivity costs, were properly introduced. These are the costs that typically separate the societal perspective from the more narrow perspectives in economic evaluations. In short, there were two methodological issues associated with both unrelated and productivity costs. Firstly, there is no consensus as to whether any of these cost items should in fact be included in the analysis. Secondly, if one decides to include them, there is no consensus as to how to estimate them.

However, hopefully these remarks on costs can be informative when the county tries to register data to create the cost weights for the mRS categories in the simulation. Once all the weights are in place, the simulations can unfold and the question of whether the concerned intervention is worthwhile can hopefully be resolved.

Bibliography

Audebert, H., Boy, S., Jankovits, R., Pilz, P., Klucken, J., Fehm, N. and Schenkel, J. (2008). Is Mobile Teleconsulting Equivalent to Hospital-Based Telestroke Services?. *Stroke*, 39(12), pp.3427-3430.

Banks, J. and Marotta, C. (2007). Outcomes Validity and Reliability of the Modified Rankin Scale: Implications for Stroke Clinical Trials: A Literature Review and Synthesis. *Stroke*, 38(3), pp.1091-1096.

Bass, E., Steinberg, E., Pitt, H., Griffiths, R., Lillemoe, K., Saba, G. and Johns, C. (1994). Comparison of the Rating Scale and the Standard Gamble in Measuring Patient Preferences for Outcomes of Gallstone Disease. *Medical Decision Making*, 14(4), pp.307-314.

Berglund, G., Abrahamsson, H. and Wilhelmsson, J. (2006). *Internmedicin*. Stockholm: Liber.

Bonita, R. and Beaglehole, R. (1988). Recovery of motor function after stroke. *Stroke*, 19(12), pp.1497-1500.

Bowry, R., Parker, S., Rajan, S., Yamal, J., Wu, T., Richardson, L., Noser, E., Persse, D., Jackson, K. and Grotta, J. (2015). Benefits of Stroke Treatment Using a Mobile Stroke Unit Compared With Standard Management. *Stroke*, 46(12), pp.3370-3374.

Brazier, J., Akehurst, R., Brennan, A., Dolan, P., Claxton, K., McCabe, C., Sculpher, M. and Tsuchyia, A. (2005). Should Patients Have a Greater Role in Valuing Health States?. *Applied Health Economics and Health Policy*, 4(4), pp.201-208.

Briggs, A., Claxton, K. and Sculpher, M. (2006). *Decision modelling for health economic evaluation*. Oxford: Oxford University Press.

Burström, K., Sun, S., Gerdtham, U., Henriksson, M., Johannesson, M., Levin, L. and Zethraeus, N. (2013). Swedish experience-based value sets for EQ-5D health states. *Qual Life Res*, 23(2), pp.431-442.

Daly, E., Roche, M., Barlow, D., Gray, A., McPherson, K., and Vessey, M. (1992). HRT: an analysis of benefits, risks and costs. *British Medical Bulletin*, 48(2), pp.368-400.

Dawson, J., Lees, J., Chang, T., Walters, M., Ali, M., Davis, S., Diener, H. and Lees, K. (2007). Association Between Disability Measures and Healthcare Costs After Initial Treatment for Acute Stroke. *Stroke*, 38(6), pp.1893-1898.

de Haan, R., Limburg, M., Bossuyt, P., van der Meulen, J. and Aaronson, N. (1995). The Clinical Meaning of Rankin `Handicap' Grades After Stroke. *Stroke*, 26(11), pp.2027-2030.

de Wit, G., Busschbach, J. and De Charro, F. (2000). Sensitivity and perspective in the valuation of health status: whose values count?. *Health Econ.*, 9(2), pp.109-126.

Dennis, M. (2000). Emotional outcomes after stroke: factors associated with poor outcome. *Journal of Neurology, Neurosurgery & Psychiatry*, 68(1), pp.47-52.

Dolan, P. (2008). Developing methods that really do value the 'Q' in the QALY. *Health Economics, Policy and Law*, 3(01).

Dolan, P. (2010). Thinking about it: thoughts about health and valuing QALYs. *Health Econ.*, 20(12), pp.1407-1416.

Dolan, P. and Kahneman, D. (2007). Interpretations Of Utility And Their Implications For The Valuation Of Health*. *The Economic Journal*, 118(525), pp.215-234.

Dolan, P., Gudex, C., Kind, P. and Williams, A. (1996). Valuing health states: A comparison of methods. *Journal of Health Economics*, 15(2), pp.209-231.

Drummond, M. (2015). *Methods for the Economic Evaluation of Health Care Programmes*. Oxford University Press.

Drummond, M., McGuire, A. and Fletcher, A. (1993). *Economic evaluation of drug therapy for hypercholesterolaemia in the United Kingdom*. University of York Centre for Health Economics.

Ebinger, M., Winter, B., Wendt, M., Weber, J., Waldschmidt, C., Rozanski, M., Kunz, A., Koch, P., Kellner, P., Gierhake, D., Villringer, K., Fiebach, J., Grittner, U., Hartmann, A., Mackert, B., Endres, M. and Audebert, H. (2014). Effect of the Use of Ambulance-Based Thrombolysis on Time to Thrombolysis in Acute Ischemic Stroke. *JAMA*, 311(16), p.1622.

Fayers, P. and Bottomley, A. (2002). Quality of life research within the EORTC—the EORTC QLQ-C30. *European Journal of Cancer*, 38, pp.125-133.

Ganesalingam, J., Pizzo, E., Morris, S., Sunderland, T., Ames, D. and Lobotesis, K. (2015). Cost-Utility Analysis of Mechanical Thrombectomy Using Stent Retrievers in Acute Ischemic Stroke. *Stroke*, 46(9), pp.2591-2598.

Ghatnekar, O. (2013). *The burden of stroke in Sweden*. Umeå: Umeå university.

Gold, M. (1996). *Cost-effectiveness in health and medicine*. New York: Oxford University Press.

Horner, R., Swanson, J., Bosworth, H. and Matchar, D. (2003). Effects of Race and Poverty on the Process and Outcome of Inpatient Rehabilitation Services Among Stroke Patients. *Stroke*, 34(4), pp.1027-1031.

Jones, P., Quirk, F., Baveystock, C. and Littlejohns, P. (1992). A Self-complete Measure of Health Status for Chronic Airflow Limitation: The St. George's Respiratory Questionnaire. *Am Rev Respir Dis*, 145(6), pp.1321-1327.

Klarman, H., Francis, J. and Rosenthal, G. (1968). Cost Effectiveness Analysis Applied to the Treatment of Chronic Renal Disease. *Medical Care*, 6(1), pp.48-54.

Lai, S., Studenski, S., Duncan, P. and Perera, S. (2002). Persisting Consequences of Stroke Measured by the Stroke Impact Scale. *Stroke*, 33(7), pp.1840-1844.

Lalonde, L., O'Connor, A., Joseph, L. and Grover, S. (2004). Health-related quality of life in cardiac patients with dyslipidemia and hypertension. *Qual Life Res*, 13(4), pp.793-804.

LaMonte, M., Xiao, Y., Hu, P., Gagliano, D., Bahouth, M., Gunawardane, R., MacKenzie, C., Gaasch, W. and Cullen, J. (2004). Shortening time to stroke treatment using ambulance telemedicine: TeleBAT. *Journal of Stroke and Cerebrovascular Diseases*, 13(4), pp.148-154.

Leppert, M., Campbell, J., Simpson, J. and Burke, J. (2015). Cost-Effectiveness of Intra-Arterial Treatment as an Adjunct to Intravenous Tissue-Type Plasminogen Activator for Acute Ischemic Stroke. *Stroke*, 46(7), pp.1870-1876.

Mackenzie, A. and Chang, A. (2002). Predictors of quality of life following stroke. *Disability and Rehabilitation*, 24(5), pp.259-265.

Mann, R., Brazier, J. and Tsuchiya, A. (2009). A comparison of patient and general population weightings of EQ-5D dimensions. *Health Econ.*, 18(3), pp.363-372.

Murray, C. and Lopez, A. (1996). *The global burden of disease*. [Cambridge, Mass.]: Published by the Harvard School of Public Health on behalf of the World Health Organization and the World Bank.

Nichols-Larsen, D., Clark, P., Zeringue, A., Greenspan, A. and Blanton, S. (2005). Factors Influencing Stroke Survivors' Quality of Life During Subacute Recovery. *Stroke*, 36(7), pp.1480-1484.

Norrving, B. (2016). *Riksstrokes årsrapport 2015*. [online] www.riksstroke.org. Available at: http://www.riksstroke.org/wp-content/uploads/2016/06/Riksstroke%C3%85rsrapport2015-PRELIMIN%C3%84R-WBB-%C3%A4ndrat-6-juli.pdf [Accessed 16 Aug. 2016].

Olai, L. (2010). *Life after a stroke event*. Uppsala: Acta Universitatis Upsaliensis.

Parker, S., Bowry, R., Wu, T., Noser, E., Jackson, K., Richardson, L., Persse, D. and Grotta, J. (2015). Establishing the First Mobile Stroke Unit in the United States. *Stroke*, 46(5), pp.1384-1391.

Rivero-Arias, O., Ouellet, M., Gray, A., Wolstenholme, J., Rothwell, P. and Luengo-Fernandez, R. (2009). Mapping the Modified Rankin Scale (mRS) Measurement into the Generic EuroQol (EQ-5D) Health Outcome. *Medical Decision Making*, 30(3), pp.341-354.

Rojas, M., Silver, A., Llewellyn, C. and Rances, L. (2005). *A Study of Adverse Occurrences and Major Functional Impairment Following Surgery*. Ft. Belvoir: Defense Technical Information Center.

Saifee, T., Burd, C. and Koppel, C. (2013). Prehospital thrombolysis in acute stroke: Results of the PHANTOM-S pilot study. *Neurology*, 81(10), pp.938-938.

Schulman, K. (1991). Cost-effectiveness of HA-1A Monoclonal Antibody for Gram-Negative Sepsis. *JAMA*, 266(24), p.3466.

Schwamm, L., Holloway, R., Amarenco, P., Audebert, H., Bakas, T., Chumbler, N., Handschu, R., Jauch, E., Knight, W., Levine, S., Mayberg, M., Meyer, B., Meyers, P., Skalabrin, E. and Wechsler, L. (2009). A Review of the Evidence for the Use of Telemedicine Within Stroke Systems of Care: A Scientific Statement From the American Heart Association/American Stroke Association. *Stroke*, 40(7), pp.2616-2634.

Slot, K., Berge, E., Dorman, P., Lewis, S., Dennis, M. and Sandercock, P. (2008). Impact of functional status at six months on long term survival in patients with ischaemic stroke: prospective cohort studies. *BMJ*, 336(7640), pp.376-379.

Smith, D., Sherriff, R., Damschroder, L., Loewenstein, G. and Ubel, P. (2006). Misremembering colostomies? Former patients give lower utility ratings than do current patients. *Health Psychology*, 25(6), pp.688-695.

Suarez-Almazor, M.E. and Connor-Spady, B. (2001). Rating of arthritis health states by patients, physicians, and the general public. Implications for cost-utility analyses. *Journal of Rheumatology*, 28(3), pp.201-216.

Ubel, P.A., Loewenstein, G. and Jepson, C. (2003). Whose quality of life? A commentary exploring discrepancy between health state evaluations of patients and the general public. *Quality of Life Research*, 12(6), pp.599-607.

Walter, S., Kostopoulos, P., Haass, A., Keller, I., Lesmeister, M., Schlechtriemen, T., Roth, C., Papanagiotou, P., Grunwald, I., Schumacher, H., Helwig, S., Viera, J., Körner, H., Alexandrou, M., Yilmaz, U., Ziegler, K., Schmidt, K., Dabew, R., Kubulus, D., Liu, Y., Volk, T., Kronfeld, K., Ruckes, C., Bertsch, T., Reith, W. and Fassbender, K. (2012). Diagnosis and treatment of patients with stroke in a mobile stroke unit versus in hospital: a randomised controlled trial. *The Lancet Neurology*, 11(5), pp.397-404.

Weinstein, M. and Manning, W. (1997). Theoretical issues in cost-effectiveness analysis. *Journal of Health Economics*, 16(1), pp.121-128.