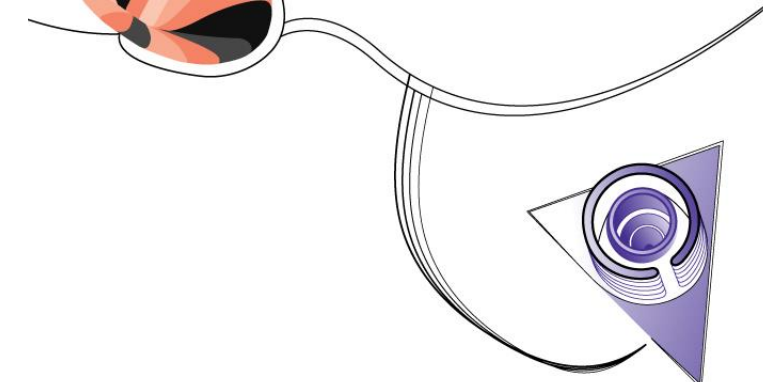
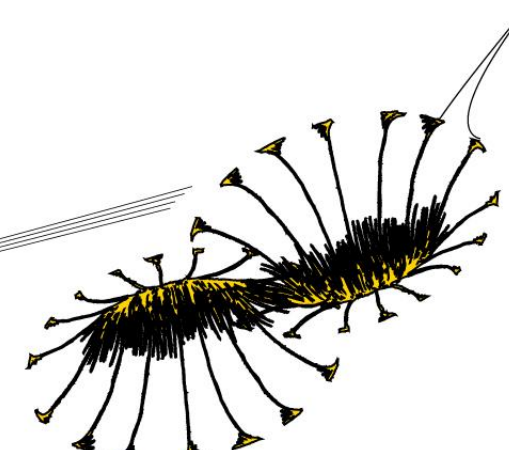


DEVELOPMENT AND VALIDATION OF A DUTCH VERSION OF THE EVIDEM FRAMEWORK



BACHELOR THESIS

Universiteit Twente

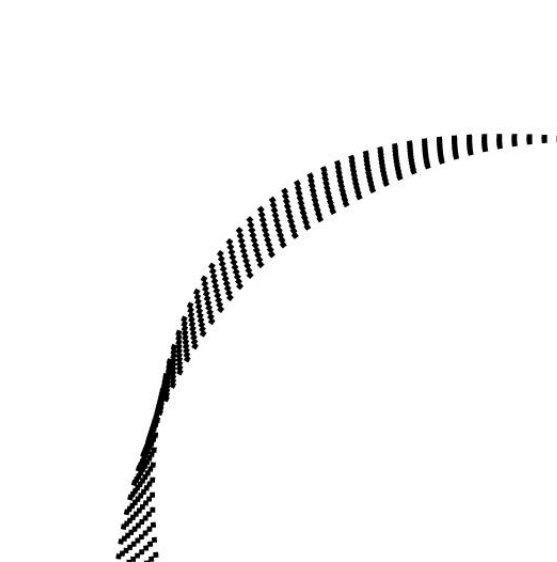
F.C. Uelderink – s1128574

17-10-2014

Begeleiders:

J.A. van Til

C.G.M. Groothuis-Oudshoorn



UNIVERSITY OF TWENTE.

Table of contents

Samenvatting	3
Summary	5
1. Introduction	7
1.1 HTA.....	7
1.2 MCDA.....	7
1.2.1 EVIDEM framework.....	7
1.3 Research questions	10
2. Method	12
2.1 Research design.....	12
2.2 Instruments	12
2.3 Research population	12
2.4 Data collection.....	13
2.5 Data analysis.....	14
3. Results	15
3.1 Response rates	15
3.2 Background characteristics.....	15
3.3 Performance of Turner Syndrome on the Decision Criteria.....	15
3.4 Intra-class coefficient	19
4. Discussion	21
4.1 Strengths and limitations	21
4.2 Recommendations	22
5. Conclusion.....	23
Bibliography.....	24
Appendix 1: Universal Criteria	25
Appendix 2: Contextual Criteria	26
Appendix 3: EVIDEM framework Version 1 (Scale 0-3).....	27
Appendix 4: EVIDEM framework Version 2 (Scale 0-100).....	34
Appendix 5: EVIDEM framework Version 3 (Scale 0-6).....	42

Samenvatting

Het is voor besluitnemers lastig om in te schatten welke interventies goed werken en kosteneffectief zijn in de gezondheidszorg. Daarom zijn er verschillende manieren om te kijken naar de verdeling van het budget voor de gezondheidszorg, waaronder Health Technology Assessment (HTA). Bij HTA is één van de onderdelen de kosteneffectiviteit, maar de problemen in de gezondheidszorg zijn breder dan alleen het kijken naar de meest kosteneffectieve interventie. Daarom wordt er vaak voorgesteld om een Multi Criteria Decision Analysis (MCDA) te gebruiken. Hierbij wordt er gekeken naar meerdere criteria van de interventie en het gezondheidsprobleem om een conclusie te kunnen trekken, maar ook met deze methode zijn er beperkingen.

Een relatief nieuw framework om beslissingen te kunnen maken over de verdeling van het budget van de gezondheidszorg aan de hand van de kosteneffectiviteit van de interventie is het EVIDEM framework. Naar het EVIDEM framework wordt nog volop onderzoek gedaan en nog niet alle voor- en nadelen van dit framework zijn bekend.

Het framework wordt momenteel gebruikt in vele talen over de wereld, dus het was in dit onderzoek van belang dat het framework eerst naar het Nederlands vertaald werd. Het doel van dit onderzoek was het vergelijken van de uitkomst van drie verschillende antwoordschalen binnen het framework. Er is gewerkt met drie verschillende antwoordschalen, waarna er is gekeken of deze schalen een verschillende uitkomst hadden en of de respondenten een voorkeur hadden voor een bepaalde schaal. Om te kijken of de hertest validiteit van het framework goed is, is er gekeken naar de hertest waardes van het framework. De respondenten zijn gevraagd of ze dezelfde vragenlijst een aantal dagen later nogmaals wilden invullen, om zo te kijken of ze dezelfde antwoorden zouden geven op dezelfde vragen.

De respondenten zijn geworven via internet en via contacten van de onderzoeker. De onderzoeker heeft zijn netwerk gevraagd om de vragenlijst in te vullen. De respondenten moesten student zijn aan de Universiteit Twente of aan de Saxion Hogeschool van Enschede, en moesten een van de volgende studies volgen: Technische Geneeskunde, Gezondheidswetenschappen, Biomedische Technologie of Verpleegkunde. Deze respondenten hebben de vragenlijst ingevuld via internet, met behulp van het software programma Lime Survey.

Aan dit onderzoek deden 38 respondenten mee, waarvan 11 hetzelfde framework nogmaals hebben ingevuld om de hertest validiteit te meten. Van de ingevulde data zijn het gemiddelde, de standaard deviatie, en het betrouwbaarheidsinterval gemeten, om de spreiding te kunnen bepalen. Met deze gegevens is vervolgens een ANOVA test uitgevoerd, om de significantie van de waarden te kunnen bepalen. Bij vijf van de 13 criteria van het framework werd een significant verschil gevonden tussen de drie antwoordschalen, op de andere acht criteria niet. Dit verschil is een indicatie dat er wel verschil

is in uitkomst tussen de 3 antwoordschalen, maar het is geen significant verschil. Alle drie de antwoordschalen hadden een ICC waarde boven de 0.05, dus de hertest van alle drie de schalen is goed. Echter hebben maar elf respondenten meegedaan aan de hertest, wat 28.9% is van de oorspronkelijke test. De meeste respondenten gaven aan dat de 0-6 schaal hun voorkeur heeft, en de huidige schaal van 0-3 werd het minst gewaardeerd door de respondenten.

Er is dus geen significant verschil in de uitkomst van de drie verschillende antwoordschalen, maar er is wel een voorkeur voor de 0-6 antwoordschaal. Alle drie de antwoordschalen hadden een goede ICC waarde, dus de hertest was voor alle drie de schalen goed. Alle drie de antwoordschalen zijn dus even goed om te gebruiken voor het framework, hoewel respondenten de voorkeur geven aan de antwoordschaal van 0 tot 6. De hertest validiteit van het framework verdient verder onderzoek.

Summary

It is difficult for decision makers in health care to assess which interventions will work well and are cost effective in healthcare. Therefore there are different ways to look at the distribution of the health budget, for example the Health Technology Assessment (HTA). In HTA, cost-effectiveness is one of the components, but the problems in health care are broader than just looking at the most cost-effective intervention. One tool that will have a broader perspective to look at is the Multi-Criteria Decision Analysis (MCDA). MCDA can check several criteria of the intervention combined with the health problem and is able to draw a conclusion based on these multiple criteria. However, also this method has its limitations.

A relatively new framework designed to make decisions about the allocation of the health budget is the EVIDEM framework. By using the EVIDEM framework the intervention gets a score for the effectiveness and the costs of the intervention, which allows the cost-effectiveness of the intervention to be compared with other interventions. Evaluation of this tool is still on going, and not all pros and cons are yet known at this time.

The framework is currently used in several languages around the world, so it was important for this research to translate the framework into Dutch. The aim of this study was to look at the impact of different response scales within the framework. Three different response scales have been used to see whether these scales have a different outcome and whether respondents have a preference for a particular scale. To investigate if the questionnaire is reliable a retest of the framework has been done as well. The respondents were asked whether they wanted to complete the same framework test a few days later, to see if they would give the same answers to the same questions.

Respondents were recruited through the Internet and through contacts of the researcher. The researcher asked his network to complete the questionnaire. The respondents needed to be a student at the University of Twente or the Saxion Highschool Enschede, and needed to attend one of the following studies: Technical Medicine, Health Sciences, Biomedical Engineering and Nursing. All respondents completed the questionnaire via the Internet, using the software program Lime Survey.

38 Respondents participated in this study, of which 11 have completed the same framework again to measure the retest validity. Based on the survey data the mean, the standard deviation and the confidence interval of the performances were measured in order to be able to determine the scatter. Based on these outcomes an ANOVA test has been performed to determine the significance of the data.

The outcome of the survey shows significant differences for five of the 13 criteria of the framework between the three response scales. For the other eight criteria no significant deviation could be found.

This difference is indication that there is a difference in performance between the three response scales, but it is not a significant difference. The retest of the framework is measured with the ICC. The ICC was for all three scales >0.05 , which means that the performance of all three scales is good. Only 11 respondents participated in the retest, which is 28.9% of the first test.

No significant difference is found in the performance of the three different response scales, but there is a preference for scale 0-6. The retest values were good for all three scales. All three response scales are equally suited to use for the framework, but the respondents have a slight preference for the response scale of 0 to 6. Further research needs to be done on the retest validity of the framework.

1. Introduction

Decision making about the appropriate allocation of healthcare resources is a necessary but complex process as it involves the consideration of a number of decision criteria (1). Pertaining health needs and accelerating technological development put an ever-increasing demand on limited health budgets. Policy makers need to make important decisions on the use of public funds. They have to choose an intervention that maximizes general population health, to reduce health inequalities of disadvantaged or vulnerable groups, and/or to respond to life-threatening situations, all with respect to practical and budgetary constraints. There are several models to aid decision making, for example: Health Technology Assessment (HTA) and Multi Criteria Decision Analysis (MCDA). The EVIDEM framework is based on the MCDA tool. This methods will be explained in chapter 1.1, 1.2 and 1.2.1.

1.1 HTA

One tool to aid decision making about the allocation of healthcare recourses is Health Technology Assessment (HTA). HTA acts as a bridge between evidence and decision-making to ensure better synthesis, communication and dissemination of information. Cost effectiveness analysis (CEA) is a part of HTA. This method currently dominates many health care policy decision-making processes. A limitation of the CEA is that it fails to address broader societal and political issues, such as disease severity, availability of alternatives, equity, and budget impact (2). Furthermore, some economic models are so complicated that they are not understood by the public and even by some decision makers (2).

1.2 MCDA

Multi Criteria Decision Analysis (MCDA) goes beyond HTA/CEA by allowing systematic and explicit consideration of multiple factors that may impact the decision. Each criterion is weighed and the performance of each healthcare intervention with respect to each criterion is scored to show an identification of weaknesses and strengths of the intervention. By giving an importance rating to all options in health, the multi-criteria decision analysis (MCDA) proves to be a method that enables people to make a considered decision of the various options.

1.2.1 EVIDEM framework

The healthcare allocation is one of the most important issues in health system research at the moment. There is a growing need for a model that systematically displays all the important factors involved in the decision making process into a model and which is also transparent and consistent(1). As seen above in Chapter 1.1 and 1.2 there are a lot of ways to weigh the decision criteria, and a relatively new method is the EVIDEM framework. The EVIDEM framework was developed by combining a multi-criteria decision analysis (MCDA) and a value matrix (VM)(3). The EVIDEM framework is a pragmatic decision-making and priority setting framework bridging health technology assessment

(HTA) and multi criteria decision analysis (MCDA). Therefore the use of the EVIDEM framework can be seen as socially relevant. (1)

The EVIDEM framework aims to consider all aspects that impact the decision making process. The framework supports consistent deliberative processes, shares decisions transparently and ensures the ranking and prioritization of interventions based on their contextual value.

There are often conflicts of interest in health care. To overcome these conflicts the costs and benefits of an intervention can be balanced against each other to evaluate a new treatment method. The EVIDEM framework will make it easy to consider the different factors and interests by following several steps. The first step is to establish the context in which the decision has to be made. Once this has been determined, it is time to look at the various options which have to be measured. For the selection of a method the objectives and criteria should contain all the options. Then the options can be scored (4). Every criterion will be scored by every respondent. A list of options is given to each respondent, whereby the most preferred option is mentioned next to the least preferred option (3). The respondents will also give a weight to every criterion. Not all criteria in the list add an equal weight to the decision which means that these options will not achieve the objective of the questionnaire in the same extent. Therefore by giving a weighting factor to all options in health, it will make sure a considered decision can be taken. After this, the results can be reviewed and a sensitivity analysis can be made (4).

The decision criteria of the EVIDEM framework consists of two parts; the normative universal criteria (MCDA Core Model) and the contextual criteria (Contextual tool) (4).

The MCDA model consist of universal criteria as the severity of the disease, the commonness of the disease, the look for unmet needs, etc. (4). The full list of universal criteria is shown in Appendix 1.

The Contextual Tool is used as a guide to tailor the framework to the context of decision-making. It includes six generic criteria/themes, with a number of sub criteria from which end-users can select the most relevant for their purpose. The Contextual Tool includes normative and feasibility criteria, such as the utility, the framework and the efficiency of the intervention (4). The full list of contextual criteria is shown in Appendix 2.

Another advantage of the EVIDEM framework is that it can be used for considering the full extent or only some aspects of it, depending on the needs. The parts of the framework are (4):

- Adapting the framework to your context, by using the Contextual Tool and assigning importance to criteria.
- Synthesizing evidence “by criterion” for the healthcare interventions to be evaluated/ ranked.

- Appraising interventions by attributing scores to each criterion of the MCDA Core Model and qualitative impacts of contextual criterion.
- Deliberating, priority-setting and decision making using a transparent and consistent approach.

Due to the fact that the EVIDEM framework is a relatively new method a lot of research still needs to be done. Most of the research focusses on the validity of the framework. In general the research outcomes show positive results in using the framework because it provides good criteria to judge optimally the innovation. As respondents indicate the assessment provides you with new insights, and therefore the framework allows you a quick optimum reflection on the performance of the innovation. Panellists indicated that the framework was a useful approach to systematic consideration of all aspects of decision, facilitating consistency, transparency and clarity of appraisal and decision-making. The strength of the framework lies in the acknowledgment and incorporation into its application that decision-making is a fundamentally value-loading enterprise. These features are combined with firm grounding in scientific evidence, which includes rigorous synthesis and quality assessment, to make the committees deliberative process as well-informed, comprehensive and explicit as possible (2).

The use of this framework also shows some disadvantages. Some respondents voiced concerns about the amount of work involved in developing the by-criterion HTA report. Others questioned the complexity of the tool and terminology used. Repeated use may be required to fully capture the utility of the framework and get acquainted with its application. Other respondents found the interpretation and utility of the MCDA value estimate challenging. Although basic principles are explicit, interpretation of the MCDA scale requires acquaintance with the broad range of criteria that are incorporated in a single number (2). The interpretation of the scale used for the MCDA decision analysis also came up for discussion. The agreement of MCDA value estimates on individual level (test-retest) is good, but large variations across panellists reflect different perspectives and personal values. For some panellists, integrating both the data and the framework in one panel session was challenging. A direct method for weight elicitation is used, which may not capture implicit or unconscious thoughts or preferences (5). Doubts can be found for the consistency of the results of the EVIDEM when comparing competing interventions, particularly when setting priorities across broad healthcare service areas. The first reason is that the EVIDEM framework ignores the contextual nature of priority setting process by assuming a set of universal priority setting criteria. The second reason is that the EVIDEM framework is vulnerable to interventions ranking inconsistency where performance evaluation of a broad range of competing interventions is mandated. These reasons lead to the question whether the approach of EVIDEM is locally meaningful and consistent when priorities are set for a range of interventions (6).

A challenge for the framework is its adaption by decision-making bodies; this will only happen if the new process is perceived as facilitating and will allow them to simplify their task, rather than only

adding complexity. Another challenge will be to bring data producers and those who make decisions together (2).

The EVIDEM Collaboration is an independent non-profit organization ran by an international Board of Directors. Its goal is to promote public health by developing efficient MCDA-based solutions to healthcare decision-making and priority setting (4). They want to overcome the above mentioned challenges by supporting and improving the use of efficient decision support theories in the continuing discussion about priorities in the health care industry. By providing free access to the decision making framework and a discussion forum on MCDA and HTA they hope to stimulate the use of the framework itself. The tool is regularly updated based on feedback from users and research teams from all over the world. When users encounter problems they can contact the organization for direct support (4). Based on all the feedback and discussions the organization is continually developing and adapting the EVIDEM framework.

By allocating a health care intervention, a lot of response scales can be used. The EVIDEM framework currently works with a 4 points Likert scale (0-3 scale). Other options that can be used are a visual analogue scale, binary outcomes, verbal descriptors or Likert scales with more or less than 4 points. Earlier research to the usability of the response scales showed different results. In this research the visual analogue scale and the Likert scales are compared for between-subject variability, consistency of relations between adjectives and scale ratings, test-retest variability, strength of correlations with other measures, performance on factor analysis and achieving a uniform distribution of responses (7). The currently used 4 point Likert scale of the EVIDEM framework gives the respondent a limited choice of answers. Because the EVIDEM framework is a relatively new method that is still in development, it is interesting to examine whether there are differences in the results for the use of different response scales.

1.3 Research questions

Further development of the EVIDEM framework is necessary before the EVIDEM framework can be applied in the Dutch health care industry. This study will focus on a few aspects of the validation of the framework, which will be discussed on the basis of the following research questions.

The main question of this investigation is ‘What is the impact of the response scale on the valuation of the performance you get in a research by students to assess a healthcare intervention?’

The main question is answered by using the following sub-questions:

- What is the difference in usability between the three response scales and which scale is preferred by the respondent?
- What is the test-retest validity of the Dutch version of the EVIDEM framework?

The purpose of this research is to translate the English EVIDEM framework to Dutch and to do further research into the development of the framework and the required adjustments for using it in the Dutch health care industry.

2. Method

2.1 Research design

The survey to validate the Dutch version of the EVIDEM framework took place in a period of ten weeks. At first a literature study was done. The literature study focussed on finding relevant information about the EVIDEM framework, as well as on other methods that are used to take decisions over healthcare resource allocation. It looked for an overview of what other methods are in place and their advantages and limitations. Also the study enabled a check of issues for which the use of the EVIDEM framework can be an improvement. After finalizing the literature study and the determination of the survey approach, the questionnaire was made and the gathering of results was started. The main question of this survey is answered by a descriptive survey.

2.2 Instruments

This questionnaire is a translation of the Canadian EVIDEM framework list into Dutch. The amounts in the questionnaire have been converted. The translation of the questionnaire had to be done with utmost care. That is why the translation of the questionnaire is checked by a person with good English language skills.

The questionnaire is used with the Turner syndrome as hypothetical subject. The Turner syndrome has been chosen as hypothetic subject because this topic is also used in Canada while testing the EVIDEM framework. This ensures that the results of the studies can be properly compared and the differences in the results are not caused by a different hypothetical situation, which makes the respondents equally well informed in both studies when completing the questionnaire. It is also expected that the population which will complete the questionnaire have affection with the topic of growth hormones, so they can empathize with the subjects.

2.3 Research population

The survey is held with a population of 38 participants. These 38 persons were divided in three subgroups. The subgroups needed to be able to compare the different scales. The groups were created randomly in Lime Survey, which made this a randomized survey. Each group used two response scales during the survey. The different response scales were:

- Method 1: 0-3 (4 points scale)
- Method 2: 0-100 (visual scale)
- Method 3: 0-6 (7 points scale)

The full questionnaire of the three methods is shown in Appendix 3, 4 and 5. For all three methods the criteria 5, 6 and 7 could also be negative. These criteria ranged from -10 to 10.

Survey group 1 received the response scales 1 and 2, survey group 2 used the scales 2 and 3, and survey group 3 worked with scales 1 and 3. This was the best method to ensure that there were nearly as much respondents in every group, which meant that the responses could be compared in good order.

The survey population had to meet with a number of in- and exclusion criteria. All people had to be a student of the University of Twente or the Saxion Highschool Enschede. These students had to follow a study in medical affairs. In practice the following studies were included: Health sciences, Technical medicine, Biomedical technology and Nursing. The people needed to study medical affairs because some in-depth knowledge of terms was required to complete the framework. It was also important to ask the age of the respondents to verify that the respondents age was a logical age for a student. The gender was not relevant.

2.4 Data collection

The questionnaires are completed on the computer. No extensive explanations were needed for working with the questionnaire, which is why it could be done by using the internet. The questionnaires could not be distributed and completed on paper, because the 0-100 scale used a slider to be dragged to the desired location, which would have been difficult to be measured on paper, because then no exact value could be picked.

The questionnaire was created with the software program Lime Survey. This program enabled the students to complete the questionnaire online. The questionnaires were distributed through internet and social media. In order to get enough respondents, reminders were sent several times.

Once a student started working with the questionnaire he or she was randomly divided into one of the subgroups. He or she completed the questionnaire with two different answering scales. At the beginning of the questionnaire the student answered the question if he or she was willing to complete the same questionnaire again, with the same answering scales as in the first survey. This follow-up study was required to measure the retest validity. That means that if the test will be done for the second time, it will check if the participants will give the same answer or a different one. If the answer will differ for both tests it means that the test-retest validity is very low and the test won't be reliable. For both tests the student had to provide his personal details; otherwise the results of the two tests couldn't be linked. The respondent was asked to add his email address, so that for the second test, a link could be sent with the same version of the questionnaire as was used the first time. When completing the second questionnaire, the respondent was asked to fill in his name or email address again to enable the linking of the tests.

Apart from that the personal information from the respondent wasn't used for any other purpose, which made this an anonymous survey. At the end a feedback questionnaire had to be answered in which the student could provide his comments for the EVIDEM framework or the answering scales.

2.5 Data analysis

After gathering the results the analysis was started. The validity of the EVIDEM framework was analysed with help of the software program SPSS. In SPSS all data from the survey was entered and the following calculations were done:

- To answer the main question of this survey the answers of both scales were converted to 0-10, to enable easier comparison of the difference scales.
- The mean, the standard deviations (SD) and the confidence interval (CI) of the performances (table 2). For all three scales the averages and the confidence interval were calculated for the performances. For the weights the average for all three scales combined together was calculated. This means that criteria 1 to 13 in table 2 are un-weighted criteria.

An ANOVA test was done for these results to enable drawing a conclusion about the significant difference of the three response scales. The ANOVA test provided a p-value. A p-value beneath 0.05 meant that there was a significant difference between the outcome of the three scales.

- The performances of these three scales for all 13 criteria were combined with the weighing of the respondent, which led to a total sum of the weighed criteria (table 2). This could be used to conclude if there was a difference in the total weighed score between the three groups.
- The Intra-class correlation coefficient (ICC) (table 5) to test the retest validity of the EVIDEM framework. The ICC is the proportion of true variance in relation to the total variance. A low ICC refers to variety in answers between the tests. The closer the values of the ICC resembled 1, the better the retest validity is. A ICC value above 0.5 is good, and a ICC value > 0.8 is very reliable. The ICC was measured per each scale for all 13 criteria together. These results of the correlation are also shown in a graph (graph 2, 3, and 4).

3. Results

3.1 Response rates

The questionnaire was sent tot 122 respondents. 84 respondents started the questionnaire but did not complete it. The analysis is bases on the 38 respondents that fully completed the questionnaire.

The attendance of the retest variability was low. 18 Respondents had left their email address for a second approach. Of these 18 respondents, 11 have completed the questionnaire for the second time. This is 28.9% of the total of all the completed questionnaires in the first test.

3.2 Background characteristics

The respondents were randomly divided into three groups, by using the software programme Lime Survey. In table 1 the background characteristics of the research population can be found. The average age of the respondents is 22 years.

Table 1; Background Characteristics

Background characteristics		
Age, Mean (years) (std)		22 (1,8)
Study, n=38, n (%)		
	Health Sciences	18 (47,3)
	Technical Medicine	11 (28,9)
	Biomedical Techonology	8 (21,1)
	Nursing	1 (2,6)
Groups (first approach), n=38, n (%)		
	Group 1	15 (39,5)
	Group 2	15 (39,5)
	Group 3	8 (21,0)
Groups (second approach), n= 11, n (%)		
	Group 1	3 (27,3)
	Group 2	5 (45,5)
	Group 3	3 (27,3)

3.3 Performance of Turner Syndrome on the Decision Criteria

The mean, the SD and the CI of the performances of the three scales are given in table 2. It can be seen that there is a spread between the individual responses of the respondents on the same scale, and there is a spread between the three scales on some criteria.

Criteria five (improvement of efficacy/ effectiveness) has a p-value of 0.000. The mean performance is very different for the scales on this criteria. The mean performance ranges from 1.3 (at scale 0-6) to

4.6 (at scale 0-3). This criteria has a significant difference between the three scales. The confidence interval shows that there is a lot of spread between the individual responses and between the scales here, for scale 0-3 is the CI 3.8 – 5.8, for scale 0-100 is the CI 3.0 – 4.8 and for scale 0-6 ranges the CI from -0.5 to 3.0. Criteria six (improvement of safety) ranges from -3.3 to -0.3 at the mean performance and has a p-value of 0.035. Criteria 8 (type of medical service) ranges two points between the 0-3/0-100 scale and the 0-6 scale. At this criteria scale 0-3 has a mean performance of 4.1, scale 0-100 of 4.6, and scale 0-6 of 6.2, with a p-value of 0.000. Criteria 12 (completeness and consistency of reporting evidence) has a p-value of 0.007 and criteria 13 (relevance and validity of evidence) has a p-value of 0.000.

The largest spread of the confidence interval is by criteria 5, 6 and 7. These are the criteria that can be negative. They have a spread up to 3.9 difference at the confidence interval. The criteria 3, 8 and 12 have the smallest spread for all three scales.

The total sum of weighted scores is highest by scale 0-6 (26.0) and lowest by scale 0-3 (22.9), see table 2.

Table 2; Descriptive Statistics of the first approach

Descriptive statistics		
	Criteria, mean (CI)	p-value
Criteria 1; Disease severity, mean (CI)		,946
Mean weight: 0,48		
Scale 0-3	6,9 (6,3 – 7,5)	
Scale 0-100	7,0 (6,3 – 7,7)	
Scale 0-6	7,0 (6,3 – 7,8)	
Criteria 2; Size of affected population, mean (CI)		,462
Mean weight: 0,57		
Scale 0-3	3,6 (2,8 – 4,3)	
Scale 0-100	2,9 (2,1 – 3,7)	
Scale 0-6	3,2 (2,4 – 4,0)	
Criteria 3; Clinical practice guidelines, mean (CI)		,746
Mean weight: 0,53		
Scale 0-3	6,9 (6,0 – 7,8)	
Scale 0-100	7,3 (6,6 – 7,9)	
Scale 0-6	7,2 (6,6 – 7,7)	
Criteria 4; Comparative interventions limitations (unmet needs), mean (CI)		,207
Mean weight: 0,57		
Scale 0-3	7,6 (6,6 – 8,5)	
Scale 0-100	6,4 (5,1 – 7,8)	
Scale 0-6	6,3 (5,0 – 7,6)	
Criteria 5; Improvement of efficacy/ effectiveness, mean (CI)		,000
Mean weight: 0,54		
Scale 0-3	4,6 (3,8 – 5,3)	

	Scale 0-100	3,9 (3,0 – 4,8)	
	Scale 0-6	1,3 (-0,5 – 3,0)	
Criteria 6; Improvement of safety, mean (CI)			,035
Mean weight: 0,51			
	Scale 0-3	-2,4 (-3,9 – -1,0)	
	Scale 0-100	-0,3 (-2,0 – 1,5)	
	Scale 0-6	-3,3 (-5,3 – -1,4)	
Criteria 7; Improvement of patient-perceived health/ PROs, mean (CI)			,934
Mean weight: 0,42			
	Scale 0-3	-0,5 (-1,2 – 0,3)	
	Scale 0-100	-0,4 (-2,2 – 1,4)	
	Scale 0-6	-0,1 (-1,3 – 1,0)	
Criteria 8; Type of medical service (therapeutic), mean (CI)			,000
Mean weight: 0,49			
	Scale 0-3	4,1 (3,3 – 4,9)	
	Scale 0-100	4,6 (3,9 – 5,3)	
	Scale 0-6	6,2 (5,6 – 6,7)	
Criteria 9; Budget impact/ cost of intervention, mean (CI)			,678
Mean weight: 0,50			
	Scale 0-3	3,0 (2,1 – 3,9)	
	Scale 0-100	3,3 (2,2 – 4,4)	
	Scale 0-6	3,6 (2,5 – 4,7)	
Criteria 10; Cost-effectiveness of intervention, mean (CI)			,253
Mean weight: 0,46			
	Scale 0-3	2,0 (1,2 – 2,8)	
	Scale 0-100	2,9 (1,9 – 3,8)	
	Scale 0-6	2,9 (1,9 – 3,9)	
Criteria 11; Impact on other spending, mean (CI)			,200
Mean weight: 0,60			
	Scale 0-3	4,7 (3,9 – 5,4)	
	Scale 0-100	5,1 (4,1 – 6,1)	
	Scale 0-6	5,8 (4,8 – 6,8)	
Criteria 12; Completeness and consistency of reporting evidence, mean (CI)			,007
Mean weight: 0,56			
	Scale 0-3	0,6 (0,1 – 1,0)	
	Scale 0-100	2,0 (1,2 – 2,8)	
	Scale 0-6	1,5 (0,7 – 2,3)	
Criteria 13; Relevance and validity of evidence, mean (CI)			,000
Mean weight: 0,59			
	Scale 0-3	2,0 (1,2 – 2,8)	
	Scale 0-100	3,7 (2,7 – 4,8)	
	Scale 0-6	5,9 (4,6 – 7,3)	
Total sum of weighted scores, Sum (std)			
	Scale 0-3	22,9 (1,5)	
	Scale 0-100	25,7 (1,3)	
	Scale 0-6	26,0 (1,6)	

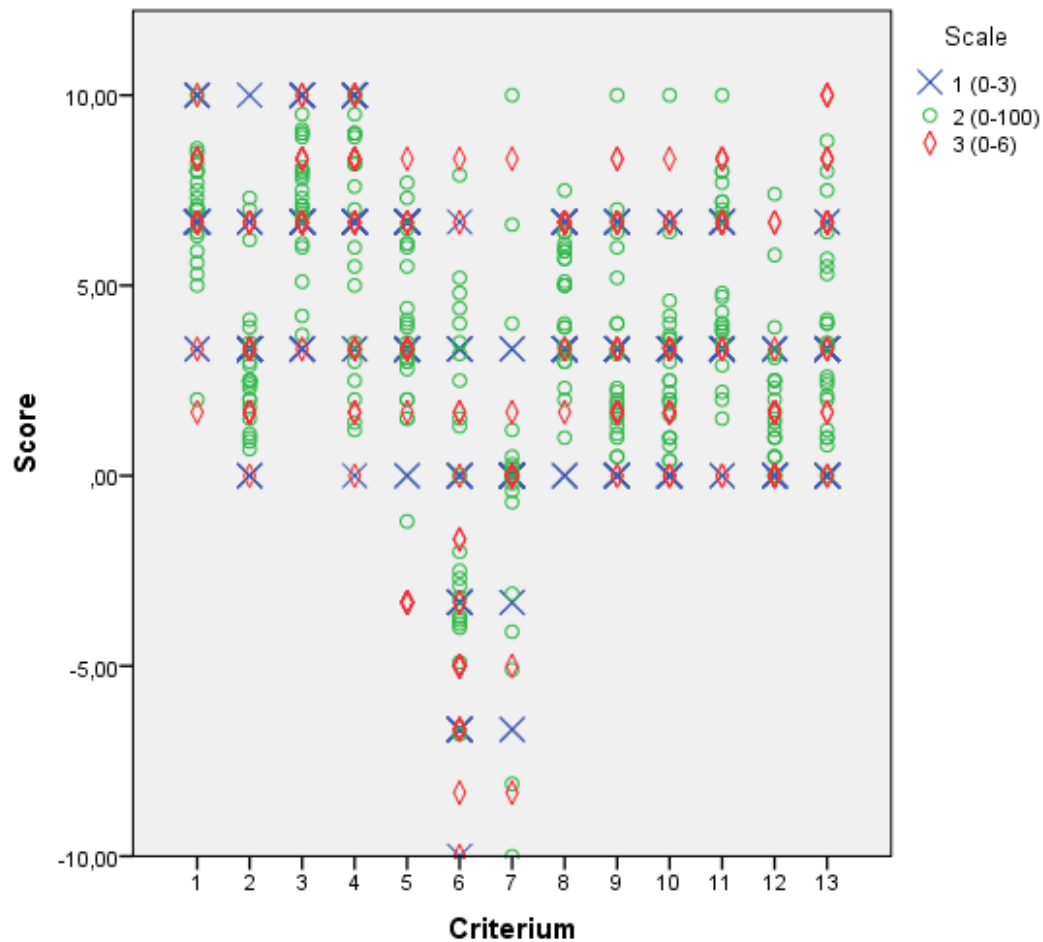


Figure 1; Performance for the criteria on three scales

Figure 1 shows the performance of the 13 criteria on the three scales. The 0-3 scale has 4 answers the respondents have to choose of. This graph shows that the respondents also like to choose the values between the points of the 0-3 scale. The respondents choose points between the values of the 0-3 scale by using the 0-6 scale and the 0-100 scale.

The results of the preferred scales by the respondents are shown in table 3. The response scale 0-6 is most preferred by respondents. Then the 0-100 scale is chosen as second preferred, and the current scale of the EVIDEM framework - the 0-3 scale - is the least preferred by the respondents. The 0-6 scale was preferred by 24 of the 38 respondents (63.2%), then the 0-100 scale followed with a large difference of seven preference votes and the 0-3 scale follows with five votes. Two respondents indicated they had no preference.

Table 3; Scale of preference

Scales of preference	
Scale, Frequency (%)	
0-3	5 (13,2)
0-100	7 (18,4)
0-6	24 (63,2)
No preference	2 (5,3)

Table 4; Table of frequency crosstab

Table of frequency			
Scale	Scale 1 (0-3)	Scale 2 (0-100)	Scale 3 (0-6)
Scale 1 (0-3)	-	5	11
Scale 2 (0-100)	0	-	5
Scale 3 (0-6)	3	1	-

There were also respondents who chose the scale they did not use themselves. For example, there were eight respondents who indicated that the 0-6 scale would have had their preference, while they used the 0-3 and 0-100 scale. These respondents were not reflected in the crosstab (table 5) above. There were also two people who used the 0-6 and 0-100 scale, who prefer the 0-3 scale, and one respondent would have preferred the 0-100 scale used while he used the 0-3 and the 0-6 scale in his research. Two respondents had no preference for a particular response scale.

3.4 Intra-class coefficient

Table 5 shows the ICC values of the retest of the EVIDEM framework. The ICC values of scale 1, 2 and 3 are all beneath 0.8, but above 0.5. Scale 0-3 has a correlation of 0.76, scale 0-100 has a correlation of 0.61 and scale 0-6 has a correlation of 0.79. The correlation can also be shown in a graph. Figure 2, 3 and 4 show the correlation between the first and second approach of all three scales.

Table 5; Intra-class correlation coefficients (ICC's) of the scores

Intra-class correlation coefficients	
Scale	ICC
Scale 0-3	0,76
Scale 0-100	0,61
Scale 0-6	0,79

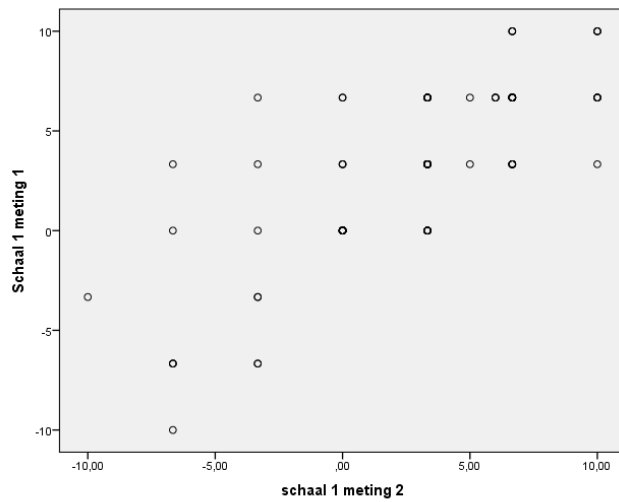


Figure 2; scatterplot scale 1

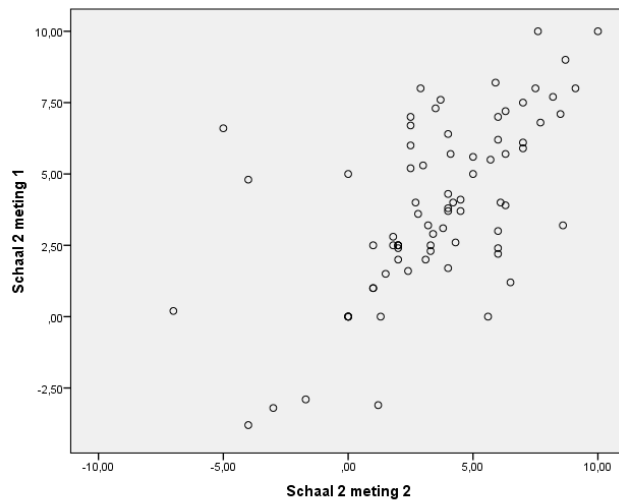


Figure 3; Scatterplot scale 2

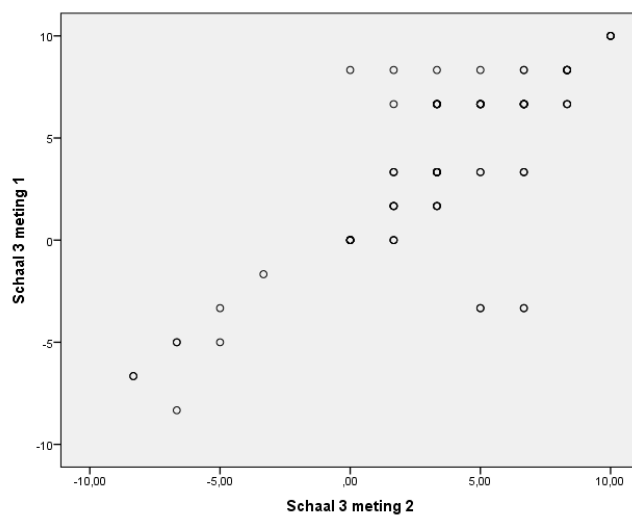


Figure 4; Scatterplot scale 3

4. Discussion

4.1 Strengths and limitations

This research has its strengths and its limitations. An advantage of this study for instance is the use of three different scales. Apart from checking if the framework is valid for the Dutch health care industry, this research also shows the preference of the respondents for a given scale.

A limitation is the number of respondents that participated in the survey. The research population needed was 60 participants, but this research only included 38 respondents. An explanation for this is the inclusion criteria that the respondents needed to be students and only in a medical study. This decreased the potential population a lot. Also the survey was sent in the beginning of the Dutch summer holiday, which meant that a lot of students were not in the mood for completing surveys anymore.

Another reason for the low participation is that this framework will be used in health care organisations that need a good framework to enable weighted decision making. The people involved in the decision making process will find it useful to complete the questionnaire as they have a direct profit from it. Students don't see this benefit and won't take time to complete the survey if there is no other benefit for them included. Mostly they are only willing to help someone by completing the questionnaire if they know the person, and when it is too long or not a topic of their interest - or too complicated - they won't complete it. This explains the high amount of missing data. The time period for the research was also short. Maybe the 60 respondents would have completed the survey if the research period would have spanned a longer time frame.

The questionnaire could only be filled in by computer, because of the 0-100 scale with a slider. The 0-100 scale can be filled in on paper, but it is easier to pick a number by using the computer. The questionnaires could therefore not be taken off on paper so the number of respondents maybe had been more when there was only a 0-3 and 0-6 scale used and questionnaires could be also filled in on paper, because then they could have been distributed on more places and could have been brought to attention to more people. Also when the questionnaires are handed out on paper the researcher can assist the respondents by completing the questionnaire. It is expected that in this case a smaller number of incomplete questionnaires would be handed in.

The groups of respondents are not all equal. The program used, Lime Survey, divides the respondents randomly into one of the three groups, but it doesn't look how many respondents are already in each group. Therefore it has happened that not all three groups include the same number of respondents which means that the preference for a scale is not entirely validated in a weighted way.

The questionnaire was long, but also complicated according to a few respondents. In the Canadian study the same (hypothetical) situation was used, but there was someone on site to provide support to the respondents and to answer their questions while they were completing the questionnaire. In this study, the questions are described in more detail, so it should be easier to understand, but some respondents found it still to be a complicated questionnaire. This may provide biased results or incomplete questionnaires by the respondent who finds it too complicated and stops completing it.

The retest validity of the framework is measured by the ICC. The ICC's of the framework were measured for 11 respondents, which is 28.9% of the first test. However, it is difficult to get a proper retest done for the third scale (the scale from 0 to 100). This scale is in fact a slider from 0 to 100, so the respondent won't most likely give exactly the same answer as in the first test. It may come close, but the intra-class correlation coefficient here indicates that there is a difference in the data. The retest validity is good because all three scales scored above 0.5 on the correlation (see table 5). However, since the retest participation is low, further research with a bigger retest population is necessary.

4.2 Recommendations

There are some limitations to this research that can be overcome by redoing the survey in another way. The first limitation is the number of respondents that took part in the study. The respondents in this study needed to be students, from four particular studies. It was difficult to find enough students that are willing to take the time to complete a long questionnaire. Therefore it is better to test this framework by allowing resources from companies operating in the healthcare to participate. They are likely to complete the questionnaire in larger numbers, because working with this framework will provide them with a benefit. These people know the topic of the questionnaire in more detail, so they will better understand it and will be more willing to spend enough time to complete the questionnaire. So a recommendation is to retest this study with resources from health care companies, to enable a bigger research population and more serious answers.

The second recommendation is based on the fact that some students found the questionnaire too complicated to complete to the end. When the Canadian researchers executed the same kind of study, experts were on site to enable sufficient assistance. This will enable higher participation and more fully completed surveys. Therefore it is recommended to let the respondents complete the questionnaire with the support of some expert of the framework and the topic. That person can give explanation to the respondents if needed, so they will also complete the questionnaire in the right way. This will prevent biased results or incomplete questionnaires by the respondent.

The last recommendation is to look at the retest of the framework. How can we make sure that more respondents will complete the questionnaire for the second time so the results will be more balanced? In this research, the retest was done by a very small group of respondents. The retest validity of the framework needs to be tested with a bigger research population.

5. Conclusion

The EVIDEM framework is a fairly new method to appreciate new health interventions. In the Netherlands, this framework has not yet been used and therefore this research has started with translating this framework in Dutch. This survey has also tested the reliability of the framework in the Dutch situation. It evaluates whether there is a difference in the answers when using different response scales and how good the retest validity of the framework is. These last questions have not been investigated before so this study has answered some new questions for using the EVIDEM framework.

The research population consisted of 38 students. The respondents have weighed the 13 criteria of the framework, and then they have scored these 13 criteria on different response scales. Each respondent was given two different response scales, but in total there are three different response scales used in the study.

Until now the EVIDEM framework has always used the 0-3 scale. In this research also two different scales have been used, the 0-6 and the 0-100 scale. The respondents of this research preferred the 0-6 scale; the current 0-3 scale was least preferred (see table 3). The total sum of weighed performances is highest for scale 0-6, and lowest for scale 0-3 (see table 2).

The mean, the SD, and the CI are measured for the 13 criteria of the EVIDEM framework. When comparing these data for the three response scales, a significant difference is only found at five of the 13 criteria, namely improvement of efficacy/ effectiveness (criteria 5), improvement of safety (criteria 6), type of medical service (criteria 8), completeness and consistency of reporting evidence (criteria 12), and relevance and validity of evidence (criteria 13) (see table 2). Five of 13 criteria showed a significant difference in performance between the three scales. So this survey has found slightly more difference than we would have expected based on chance only. This is only an indication but not a solid conclusion that there is a difference between the 3 methods. Until further investigation is done, all three response scales can be used for the EVIDEM framework.

Of the 38 respondents in the study, 11 participated in the retest study. This is a low turnout (28.9% of the first test). No difference for the three response scales is shown in the performances of the retest. All three response scales had a good validity on the retest as they all had an ICC above 0.5 (see table 5). Because of this outcome there is still need for further research on the retest of these scales.

Bibliography

1. Rob Baltussen LN. Priority setting of health interventions: the need for multi-criteria decision analysis. *Cost Effectiveness and Resource allocation*. 2006;4(14):9.
2. Michèle Tony MW, Hanane Khoury, Donna Rindress, Tina Papastavrow, Paul Oh, Mireille M Goetghebeur. Bridging health technology assessment (HTA) with multicriteria decision analyses (MCDA): field testing of the EVIDEM framework for coverage decisions by a public payer in Canada. *BMC Health Services Research*. 2011;11(329):13.
3. Mireille M Goetghebeur MW, Hanane Khoury, Randy J Levitt, Lonny Erickson, Donna Rindress. Evidence and Value: Impact on DEcisionMaking - the EVIDEM framework and potential applications. *BMC Health Services Research*. 2008;8(270):16.
4. Collaboration TE. The EVIDEM website. 2014.
5. Mireille M Goetghebeur MW, Hanane Khoury, Donna Rindress, Jean-Pierre Grégoire, Cheri Deal. Combining multicriteria decision analysis, ethics and health technology assessment: applying the EVIDEM decisionmaking framework to growth hormone for Turner syndrome patients. *Cost Effectiveness and Resource allocation*. 2010;8(4):15.
6. Sitaporn Youngkong NT, Dereck Chitama. The EVIDEM framework and its usefulness for priority setting across a broad range of health interventions; Commentary. *Cost Effectiveness and Resource allocation*. 2011;9(8):3.
7. Guyatt GH, Townsend, M., Berman, L.B., Keller, J.L. A comparison of Likert and visual analogue scales for measuring change in function. *Pergamon Journals Ltd*. 1986;40(12):1129-33.

Appendix 1: Universal Criteria

The universal criteria of the MCDA model are (4):

- The severity of the disease
- Is the disease common?
- Has the disease many unmet needs?
- Recommended in consensus guidelines by experts
- Does the intervention improve the efficacy/ effectiveness over standard of care?
- Does the intervention improve the safety and tolerability over standard of care?
- Does the intervention improve the patient-reported outcomes/ perceived health over standard of care?
- Is there a major risk reduction or a major alleviation of suffering?
- What are the savings of the treatment?
- The quality of the evidence

Appendix 2: Contextual Criteria

The contextual criteria are (4):

- Utility (Goals of the health care plan)
- Fairness
- Efficiency
- System capacity and requirements
- Pressures / barriers from stakeholders
- Political and historical context
- Environmental impact of the intervention

Appendix 3: EVIDEM framework Version 1 (Scale 0-3)

Voor u ligt een vragenlijst over een interventie voor groeihormonen bij jonge meisjes met het syndroom van Turner. Het betreft hier een hypothetische situatie.

Deze vragenlijst is bedoeld voor studenten aan de Universiteit Twente met de studies Gezondheidswetenschappen, Technische Geneeskunde of Biomedische Technologie, en studenten van de studie Verpleegkunde aan het Saxion Enschede.

In de eerste helft van de vragenlijst wordt er gevraagd hoe belangrijk u elk criterium vindt in de waardering van een gezondheidsinterventie. In het tweede deel wordt u gevraagd om aan te geven hoe u de ziekte van Turner waardeert op deze criteria.

In deze vragenlijst zullen alle vragen 2 keer voorkomen, maar met een verschillende antwoordschaal.

Deze vragenlijst bestaat uit 44 vragen en zal ongeveer een kwartier de tijd kosten.

De hypothetische situatie in dit onderzoek is het toedienen van groeihormonen (GH) bij jonge meisjes met het syndroom van Turner.

Medicijn: polypeptide hormoon

Indicatie: behandeling van een korte gestalte bij meisjes met het syndroom van Turner

Behandeling: subcutane injectie 3 tot 7 dagen per week

Duur van de interventie: Moet worden vastgesteld. Start direct als groeistoornis is geconstateerd tot een bevredigende lengte bereikt is (6 jaar behandeling vanaf 10 jaar)

Vergelijking(en): Zonder behandeling

Economische last van de ziekte: Geen gegevens beschikbaar.

Wat is uw leeftijd?

Answer

Only numbers may be entered in this field.

Wat studeert u?

Choose one of the following answers

- ☐ Technische Geneeskunde
- ☐ Biomedische Technologie
- ☐ Gezondheidswetenschappen
- ☐ Verpleegkunde

Vult u hier uw email adres of naam in, zodat we de resultaten van uw 2 ingevulde vragenlijsten aan elkaar kunnen koppelen.

Answer

geef aan hoe belangrijk je elk criterium vind in de beoordeling van een gezondheidsinterventie. Als je gelooft dat een criterium systematisch niet zou moeten worden beschouwd, ken een 0 toe.

Ernst van de ziekte

	1 (niet belangrijk)	2	3	4	5 (belangrijk)	0 (moet systematisch niet worden beschouwd)
Ernst van de ziekte	☐	☐	☐	☐	☐	☐

? De ernst van de aandoening van de patiënten behandeld met de voorgestelde interventie (of de ernst van de aandoening die voorkomen dient te worden) in termen van de sterfte, invaliditeit, impact op de kwaliteit van leven, het verloop van de ziekte (d.w.z. spoedeisend, verloop klinische fases).

Grootte van de beïnvloede populatie

	1 (niet belangrijk)	2	3	4	5 (belangrijk)	0 (moet systematisch niet worden beschouwd)
Grootte van de beïnvloede populatie:	☐	☐	☐	☐	☐	☐

? Aantal mensen getroffen door de aandoening (behandeld of voorkomen door de voorgestelde interventie) in een bepaalde populatie op een bepaald tijdstip; dit kan worden uitgedrukt als jaarlijks aantal nieuwe gevallen (jaarlijkse incidentie) en/ of deel van de getroffen bevolking op een bepaald punt in de tijd (prevalentie).

Aansluiting tussen de interventie en de klinische praktijk richtlijnen

	1 (niet belangrijk)	2	3	4	5 (belangrijk)	0 (moet systematisch niet worden beschouwd)
Klinische praktijk richtlijnen:	☐	☐	☐	☐	☐	☐

? Overeenkomst tussen de voorgestelde interventie (of van soortgelijke alternatieven) met de huidige consensus van deskundigen over wat bijdraagt aan de **state-of-the-art** voor de management van de beoogde aandoening; richtlijnen worden gewoonlijk ontwikkeld via een expliciet proces en zijn bedoeld om de klinische praktijk te verbeteren.

Beperkingen van vergelijkbare interventies (onvervulde behoeften)

	1 (niet belangrijk)	2	3	4	5 (belangrijk)	0 (moet systematisch niet worden beschouwd)
Beperkingen van vergelijkbare interventies (onvervulde behoeften)	☐	☐	☐	☐	☐	☐

? Tekortkomingen van vergelijkbare interventies in hun vermogen om de aandoening te voorkomen, genezen of de gevolgen te verminderen; ook tekortkomingen met betrekking tot veiligheid, patiënt-gerapporteerde uitkomsten en geschiktheid.

Verbetering van de werkzaamheid/ effectiviteit

	1 (niet belangrijk)	2	3	4	5 (belangrijk)	0 (moet systematisch niet worden beschouwd)
Verbetering van de werkzaamheid/effectiviteit	☐	☐	☐	☐	☐	☐

? De mogelijkheden van de voorgestelde interventie om te komen tot een gewenste (positieve) verandering in de tekenen, symptomen en het verloop van de beoogde aandoening in aanvulling op de gunstige veranderingen veroorzaakt door andere, bestaande interventies. Inclusief gegevens over werkzaamheid en de effectiviteit, indien beschikbaar.

Verbetering van de veiligheid

? Impact van het verstrekken van dekking voor de voorgestelde interventie op andere uitgaven (uitgezonderd interventie kosten), zoals ziekenhuisopname, consulten bij specialisten, bijwerkingen, langdurige zorg, kosten invaliditeit, verlies van productiviteit, tijd van de verzorger, etc.

Volledigheid en samenhang van het gerapporteerde bewijs (het voldoen aan de wetenschappelijke rapportage normen en eenduidigheid met de bronnen)

	1 (niet belangrijk)	2	3	4	5 (belangrijk)	0 (moet systematisch niet worden beschouwd)
Volledigheid en samenhang van het gerapporteerde bewijs:	☐	☐	☐	☐	☐	☐

? Mate waarin de gerapporteerde gegevens over de voorgestelde interventie compleet zijn (dat wil zeggen, het voldoen aan wetenschappelijke normen voor rapportage) en in overeenstemming zijn met de aangehaalde bronnen.

Relevantie en geldigheid van het bewijs (relevant voor de besluitnemers & voldoen aan de wetenschappelijke normen)

	1 (niet belangrijk)	2	3	4	5 (belangrijk)	0 (moet systematisch niet worden beschouwd)
Relevantie en geldigheid van het bewijs:	☐	☐	☐	☐	☐	☐

? Mate waarin het bewijs voor de voorgestelde interventie relevant is voor het besluitvormend orgaan (in termen van populatie, ziekte stadium, vergelijkbare interventies, resultaten, etc.) en in overeenstemming met de wetenschappelijke normen (d.w.z. onderzoeksopzet, etc.) en conclusies (overeenstemming van resultaten tussen de studies). Dit is inclusief het in aanmerking nemen van onzekere factoren (bijvoorbeeld tegenstrijdige resultaten tussen de onderzoeken, beperkt aantal onderzoeken en patiënten).

Let op: Dezelfde 14 vragen als hiervoor zullen nu nogmaals volgen, maar met een andere antwoordschaal.

De ziekte van Turner is een vrouw-specifieke genetische aandoening die wordt gekenmerkt door een korte gestalte, cardiovasculaire gebreken, het ontbreken van de puberteit, onvruchtbaarheid, verhoogd risico op diabetes, defecten in visueel-ruimtelijke organisatie en non-verbale probleemoplossing, psychosociale problemen en een verminderde levensverwachting.

Waardeer de ernst van deze ziekte:

	0 (niet zwaar)	1	2	3 (heel zwaar)
Ernst van de ziekte	☐	☐	☐	☐

De prevalentie is 40/100.000 vrouwelijke volwassenen.

Waardeer de grootte van de getroffen populatie:

	0 (hele zeldzame ziekte)	1	2	3 (veel voorkomende ziekte)
Grootte van de getroffen populatie	☐	☐	☐	☐

De internationale richtlijnen voor de ziekte van Turner schrijven voor dat de groeihormoon behandeling overwogen moet worden zodra de groeistoornis wordt geconstateerd en potentiële risico's/ voordelen zijn besproken met de patiënt / familie. Behandelen totdat een bevredigende lengte is bereikt.

Waardeer de aansluiting van de interventie op de bestaande klinische praktijkrichtlijnen:

	0 (niet aanbevolen)	1	2	3 (sterkte aanbeveling)
Klinische praktijkrichtlijnen	☐	☐	☐	☐

Er is geen andere therapeutische interventie geïndiceerd om korte gestalte te behandelen bij het syndroom van Turner.
Waardeer de beperkingen van vergelijkbare interventies (onvervulde behoeften):

	0 (geen of heel weinig beperkingen - onvervulde behoeften)	1	2	3 (belangrijke beperkingen - onvervulde behoeften)
Vergelijkbare interventies:	☐	☐	☐	☐

Uitkomsten van de interventie:

In 4 placebo controlled RCTs was het **verschil met onbehandelde patienten 7 cm**.
 In observationeel gecontroleerde studies was het **verschil met de controlegroep = 2,1-6,8 cm**
Waardeer de verbetering van de werkzaamheid/ effectiviteit:

	-3 (lager dan zonder behandeling)	-2	-1	0 (geen verbetering)	1	2	3 (grote verbetering t.o.v. geen behandeling)
Werkzaamheid/ effectiviteit:	☐	☐	☐	☐	☐	☐	☐

De meest voorkomende bijwerkingen van het medicijn zijn: Chirurgie (50%), oorproblemen (6% tot 47%) gewrichts- (13,5%) en respiratoire (11%) aandoeningen, sinusitis (18,9%).
Waardeer de veiligheid van het medicijn:

	-3 (veel minder veilig dan zonder behandeling)	-2	-1	0 (geen verbetering)	1	2	3 (veel veiliger dan geen behandeling)
Veiligheid:	☐	☐	☐	☐	☐	☐	☐

Er zijn geen significantie verschillen in de verbetering van de ervaren gezondheid.
Waardeer de verbetering van de ervaren gezondheid/ patiënt gerapporteerde uitkomsten:

	-3 (lager dan zonder behandeling)	-2	-1	0 (geen verbetering)	1	2	3 (grote verbetering t.o.v. geen behandeling)
De ervaren gezondheid/ patiënt gerapporteerde uitkomsten:	☐	☐	☐	☐	☐	☐	☐

Type voordeel:

Het doel van de behandeling is het stimuleren van de groei en het verbeteren van het psychosociale welzijn (lengte winst 7 cm, patiënt gerapporteerde uitkomsten gegevens beperkt en niet eenduidig).
Waardeer de aard van de medische dienst (therapeutisch):

	0 (onbelangrijke dienst)	1	2	3 (een belangrijke dienst)
Aard van de medische dienst (therapeutisch):	☐	☐	☐	☐

Economie:

Gemiddelde jaarlijkse kosten van het geneesmiddel per patiënt: €19.000

Jaarimpact voor de Nederlandse publieke medicijn budgetten: €7.700.000 (dekking voor alle patiënten)

De totale uitgaven voor geneesmiddelen zijn in Nederland €5,9 miljard per jaar.

Waardeer de consequenties voor het budget/ kosten van de interventie:

	0 (hoge bijkomende uitgaven)	1	2	3 (flinke besparingen)
Consequenties voor het budget/ kosten van de interventie:	☐	☐	☐	☐

Extra kosten per extra centimeter in de uiteindelijke lengte: € 16.000.

Extra kosten per gewonnen QALY: € 165.000.

Waardeer de kosten-effectiviteit van de interventie:

	0 (niet kosten effectief)	1	2	3 (zeer kosten effectief)
Kosteneffectiviteit van de interventie:	☐	☐	☐	☐

Extra kosten per patiënt: € 794

Deze kosten zijn niet gemaakt door de patiënt, maar dit zijn kosten die gemaakt moeten worden voor de behandeling van de patiënt, bijv. training door de verpleegkundige, poliklinische bezoeken en röntgenfoto's gedurende 6 jaar (exclusief kosten van geneesmiddelen).

Waardeer de impact van de interventie op andere uitgaven:

	0 (hoge bijkomende uitgaven)	1	2	3 (flinke besparingen)
Impact op andere uitgaven:	☐	☐	☐	☐

Kwaliteit van het bewijs:

Epidemiologie: beperkte statistische informatie;

Klinische gegevens: beperkte rapportage van bijwerkingen;

PRO (patiënt gerapporteerde uitkomsten): onvolledige rapportage van de vragenlijst aspecten;

Economische impact: sommige kenmerken van het model zijn onduidelijk;

Budgettaire gevolgen: geen sensitiviteitsanalyse gemeld;

Waardeer de compleetheid en eenduidigheid van het gerapporteerde bewijs:

	0 (veel missende informatie/ niet eenduidig)	1	2	3 (compleet en eenduidig)
Het gerapporteerde bewijs:	☐	☐	☐	☐

Epidemiologie: 4 placebo gecontroleerde RCTs (N= 42-104, 1 in Canada, 3 in de VS); Observationeel gecontroleerde studies (N= 26-123, 1 in Duitsland, 1 in Griekenland, 1 in Israël, 3 in Italië);

Klinische gegevens: onzekerheid over de uiteindelijke lengte winst, hoog verloop in belangrijke RCT;

PRO: tussentijdse analyse van een subgroup van de deelnemers aan een niet anonieme RCT;

Economische evaluatie: twijfelachtige uitkomst-kosten per cm van de uiteindelijke lengte, geen bijwerkingen kosten, zwakke bruikbaarheidsgegevens;

Invloed budget: model neemt aan dat alle meisjes in Canada worden behandeld.

Waardeer de relevantie en betrouwbaarheid van deze gegevens:

	0 (lage relevantie/ geldigheid)	1	2	3 (hoge relevantie/ geldigheid)
Het bewijs:	☐	☐	☐	☐

EIND VAN DE VRAGENLIJST

In dit onderzoek zijn drie verschillende antwoordschalen gebruikt. Hiervan heeft u er twee te zien gekregen. Heeft u een voorkeur voor een van de twee gebruikte antwoordschalen?

Choose one of the following answers

- ☐ 0-3 (4 punts schaal)
- ☐ 0-6 (7 punts schaal)
- ☐ 0-100 (visuele schaal)
- ☐ Geen voorkeur

Heeft u nog opmerkingen?

Answer

Bedankt voor het invullen van de vragenlijst! Voor vragen kunt u contact opnemen met f.c.uelderink@student.utwente.nl

Appendix 4: EVIDEM framework Version 2 (Scale 0-100)

Voor u ligt een vragenlijst over een interventie voor groeihormonen bij jonge meisjes met het syndroom van Turner. Het betreft hier een hypothetische situatie.

Deze vragenlijst is bedoeld voor studenten aan de Universiteit Twente met de studies Gezondheidswetenschappen, Technische Geneeskunde of Biomedische Technologie, en studenten van de studie Verpleegkunde aan het Saxion Enschede.

In de eerste helft van de vragenlijst wordt er gevraagd hoe belangrijk u elk criterium vindt in de waardering van een gezondheidsinterventie. In het tweede deel wordt u gevraagd om aan te geven hoe u de ziekte van Turner waardeert op deze criteria.

In deze vragenlijst zullen alle vragen 2 keer voorkomen, maar met een verschillende antwoordschaal.

Deze vragenlijst bestaat uit 44 vragen en zal ongeveer een kwartier de tijd kosten.

De hypothetische situatie in dit onderzoek is het toedienen van groeihormonen (GH) bij jonge meisjes met het syndroom van Turner.

Medicijn: polypeptide hormoon

Indicatie: behandeling van een korte gestalte bij meisjes met het syndroom van Turner

Behandeling: subcutane injectie 3 tot 7 dagen per week

Duur van de interventie: Moet worden vastgesteld. Start direct als groeistoornis is geconstateerd tot een bevredigende lengte bereikt is (6 jaar behandeling vanaf 10 jaar)

Vergelijking(en): Zonder behandeling

Economische last van de ziekte: Geen gegevens beschikbaar.

Wat is uw leeftijd?

Answer

Only numbers may be entered in this field.

Wat studeert u?

Choose one of the following answers

- ☐ Technische Geneeskunde
- ☐ Biomedische Technologie
- ☐ Gezondheidswetenschappen
- ☐ Verpleegkunde

Vult u hier uw email adres of naam in, zodat we de resultaten van uw 2 ingevulde vragenlijsten aan elkaar kunnen koppelen.

Answer

WEGING VAN DE CRITERIA

Geef aan hoe belangrijk je elk criterium vind in de beoordeling van een gezondheidsinterventie. Als je gelooft dat een criterium systematisch niet zou moeten worden beschouwd, ken een 0 toe.

Ernst van de ziekte

	1 (niet belangrijk)	2	3	4	5 (belangrijk)	0 (moet systematisch niet worden beschouwd)
Ernst van de ziekte	☐	☐	☐	☐	☐	☐

? De ernst van de aandoening van de patiënten behandeld met de voorgestelde interventie (of de ernst van de aandoening die voorkomen dient te worden) in termen van de sterfte, invaliditeit, impact op de kwaliteit van leven, het verloop van de ziekte (d.w.z. spoedeisend, verloop klinische fases).

Grootte van de beïnvloede populatie

	1 (niet belangrijk)	2	3	4	5 (belangrijk)	0 (moet systematisch niet worden beschouwd)
Grootte van de beïnvloede populatie:	☐	☐	☐	☐	☐	☐

? Aantal mensen getroffen door de aandoening (behandeld of voorkomen door de voorgestelde interventie) in een bepaalde populatie op een bepaald tijdstip; dit kan worden uitgedrukt als jaarlijks aantal nieuwe gevallen (jaarlijkse incidentie) en/ of deel van de getroffen bevolking op een bepaald punt in de tijd (prevalentie).

Aansluiting tussen de interventie en de klinische praktijk richtlijnen

	1 (niet belangrijk)	2	3	4	5 (belangrijk)	0 (moet systematisch niet worden beschouwd)
Klinische praktijk richtlijnen:	☐	☐	☐	☐	☐	☐

? Overeenkomst tussen de voorgestelde interventie (of van soortgelijke alternatieven) met de huidige consensus van deskundigen over wat bijdraagt aan de **state-of-the-art** voor de management van de beoogde aandoening; richtlijnen worden gewoonlijk ontwikkeld via een expliciet proces en zijn bedoeld om de klinische praktijk te verbeteren.

Beperkingen van vergelijkbare interventies (onvervulde behoeften)

	1 (niet belangrijk)	2	3	4	5 (belangrijk)	0 (moet systematisch niet worden beschouwd)
Beperkingen van vergelijkbare interventies (onvervulde behoeften)	☐	☐	☐	☐	☐	☐

? Tekortkomingen van vergelijkbare interventies in hun vermogen om de aandoening te voorkomen, genezen of de gevolgen te verminderen; ook tekortkomingen met betrekking tot veiligheid, patiënt-gerapporteerde uitkomsten en geschiktheid.

Verbetering van de werkzaamheid/ effectiviteit

	1 (niet belangrijk)	2	3	4	5 (belangrijk)	0 (moet systematisch niet worden beschouwd)
Verbetering van de werkzaamheid/ effectiviteit	☐	☐	☐	☐	☐	☐
? De mogelijkheden van de voorgestelde interventie om te komen tot een gewenste (positieve) verandering in de tekenen, symptomen en het verloop van de beoogde aandoening in aanvulling op de gunstige veranderingen veroorzaakt door andere, bestaande interventies. Inclusief gegevens over werkzaamheid en de effectiviteit, indien beschikbaar. Verbetering van de veiligheid						
Verbetering van de veiligheid	☐	☐	☐	☐	☐	☐
? De mogelijkheden van de voorgestelde interventie om tot een vermindering van de aan de interventie gerelateerde schadelijke of ongewenste effecten op de aandoening te komen, in vergelijking met andere interventies. Verbetering van de patiënt-ervaren gezondheid/ patiënt gerapporteerde uitkomsten						
Verbetering van de patiënt-ervaren gezondheid/ patiënt gerapporteerde uitkomsten:	☐	☐	☐	☐	☐	☐
? Capaciteit van de voorgestelde interventie om te komen tot positieve veranderingen in door de patiënt gerapporteerde uitkomsten (bijv. kwaliteit van leven) in aanvulling op gunstige veranderingen veroorzaakt door andere interventies; inclusief verbetering in het comfort van patiënten. Aard van de medische dienst (therapeutisch)						
Aard van de medische dienst (therapeutisch):	☐	☐	☐	☐	☐	☐
? Aard van het klinische voordeel veroorzaakt door de voorgestelde interventie op patiënt-niveau (bijv. verlichting van de symptomen, het verlengen van de levensduur, genezing). Consequenties voor het budget/ kosten van de interventie						
Consequenties voor het budget/ kosten van de interventie:	☐	☐	☐	☐	☐	☐
? Netto effect van de vergoeding van de interventie op het budget van de basisverzekering (exclusief overige uitgaven). Dit is het verschil tussen de verwachte uitgaven voor de voorgestelde interventie en de potentiële kostenbesparingen die kunnen voortvloeien uit de vervanging van andere interventie(s) die nu vallen onder de basisverzekering. Beperkt tot de kosten van de interventie (bijv. aanschafkosten, implementatie- en onderhoudskosten).						

Kosten-effectiviteit van de interventie

	1 (niet belangrijk)	2	3	4	5 (belangrijk)	0 (moet systematisch niet worden beschouwd)
Kosten-effectiviteit van de interventie:	☐	☐	☐	☐	☐	☐

? Verhouding van de extra kosten van de voorgestelde interventie tot het bijkomende voordeel in vergelijking met alternatieven. Voordeel kan worden uitgedrukt als het aantal gebeurtenissen die zijn vermeden, gewonnen levensjaren, verbeterde kwaliteit van de gewonnen levensjaren, extra pijn-vrije dagen, etc.

Impact op andere uitgaven (bijv. ziekenhuisopname, invaliditeit)

	1 (niet belangrijk)	2	3	4	5 (belangrijk)	0 (moet systematisch niet worden beschouwd)
Impact op andere uitgaven:	☐	☐	☐	☐	☐	☐

? Impact van het verstrekken van dekking voor de voorgestelde interventie op andere uitgaven (uitgezonderd interventie kosten), zoals ziekenhuisopname, consulten bij specialisten, bijwerkingen, langdurige zorg, kosten invaliditeit, verlies van productiviteit, tijd van de verzorger, etc.

Volledigheid en samenhang van het gerapporteerde bewijs (het voldoen aan de wetenschappelijke rapportage normen en eenduidigheid met de bronnen)

	1 (niet belangrijk)	2	3	4	5 (belangrijk)	0 (moet systematisch niet worden beschouwd)
Volledigheid en samenhang van het gerapporteerde bewijs:	☐	☐	☐	☐	☐	☐

? Mate waarin de gerapporteerde gegevens over de voorgestelde interventie compleet zijn (dat wil zeggen, het voldoen aan wetenschappelijke normen voor rapportage) en in overeenstemming zijn met de aangehaalde bronnen.

Relevantie en geldigheid van het bewijs (relevant voor de besluitnemers & voldoen aan de wetenschappelijke normen)

	1 (niet belangrijk)	2	3	4	5 (belangrijk)	0 (moet systematisch niet worden beschouwd)
Relevantie en geldigheid van het bewijs:	☐	☐	☐	☐	☐	☐

? Mate waarin het bewijs voor de voorgestelde interventie relevant is voor het besluitvormend orgaan (in termen van populatie, ziekte stadium, vergelijkbare interventies, resultaten, etc.) en in overeenstemming met de wetenschappelijke normen (d.w.z. onderzoeksopzet, etc.) en conclusies (overeenstemming van resultaten tussen de studies). Dit is inclusief het in aanmerking nemen van onzekere factoren (bijvoorbeeld tegenstrijdige resultaten tussen de onderzoeken, beperkt aantal onderzoeken en patiënten).

Let op: Dezelfde 14 vragen als hiervoor zullen nu nogmaals volgen, maar met een andere antwoordschaal.

De ziekte van Turner is een vrouwspecifieke genetische aandoening die wordt gekenmerkt door een korte gestalte, cardiovasculaire gebreken, het ontbreken van de puberteit, onvruchtbaarheid, verhoogd risico op diabetes, defecten in visueel-ruimtelijke organisatie en non-verbale probleemoplossing, psychosociale problemen en een verminderde levensverwachting.

Waardeer de ernst van deze ziekte:

Kies een getal tussen 0 (niet ernstig) en 100 (heel ernstig)

Ernst van de ziekte



De prevalentie is 40/100.000 vrouwelijke volwassenen.

Waardeer de grootte van de getroffen populatie:

Kies een getal tussen 0 (hele zeldzame ziekte) en 100 (veelvoorkomende ziekte)

Grootte van de getroffen populatie



De internationale richtlijnen voor de ziekte van Turner schrijven voor dat de groeihormoon behandeling overwogen moet worden zodra de groeistoornis wordt geconstateerd en potentiële risico's/ voordelen zijn besproken met de patiënt / familie. Behandelen totdat een bevredigende lengte is bereikt.

Waardeer de aansluiting van de interventie op de bestaande klinische praktijkrichtlijnen:

Kies een getal tussen 0 (niet aanbevolen) en 100 (sterk aanbevolen)

Klinische praktijkrichtlijnen



Er is geen andere therapeutische interventie geïndiceerd om korte gestalte te behandelen bij het syndroom van Turner.

Waardeer de beperkingen van vergelijkbare interventies (onvervulde behoeften):

Kies een getal tussen 0 (geen of heel weinig beperkingen - onvervulde behoeften) en 100 (belangrijke beperkingen - onvervulde behoeften)

Vergelijkbare interventies



Uitkomsten van de interventie:

In 4 placebo controlled RCTs was het verschil met onbehandelde patienten 7 cm.

In observationeel gecontroleerde studies was het verschil met de controlegroep = 2,1-6,8 cm.

Waardeer de verbetering van de werkzaamheid/ effectiviteit:

Kies een getal tussen -100 (lager dan zonder behandeling) en 100 (grote verbetering t.o.v. geen behandeling), waarbij 0 = geen verbetering.

Werkzaamheid/ effectiviteit



De meest voorkomende bijwerkingen van het medicijn zijn: Chirurgie (50%), oorproblemen (6% tot 47%) gewrichts- (13,5%) en respiratoire (11%) aandoeningen, sinusitis (18,9%).

Waardeer de veiligheid van het medicijn:

Kies een getal tussen -100 (veel minder veilig dan geen behandeling) en 100 (veel veiliger dan zonder behandeling), waarbij 0 = geen verbetering.

Veiligheid



Er zijn geen significantie verschillen in de verbetering van de ervaren gezondheid.

Waardeer de verbetering van de ervaren gezondheid/ patiënt gerapporteerde uitkomsten:

Kies een getal tussen -100 (lager dan geen behandeling) en 100 (grote verbetering t.o.v. geen behandeling), waarbij 0 = geen verbetering.

De ervaren gezondheid/ patiënt gerapporteerde uitkomsten



Type voordeel:

Het doel van de behandeling is het stimuleren van de groei en het verbeteren van het psychosociale welzijn (lengte winst 7 cm, patiënt gerapporteerde uitkomsten gegevens beperkt en niet eenduidig).

Waardeer de aard van de medische dienst (therapeutisch):

Kies een getal tussen 0 (onbelangrijke dienst) en 100 (een belangrijke dienst (bijvoorbeeld genezing)).

Aard van de medische dienst (therapeutisch)



Economie:

Gemiddelde jaarlijkse kosten van het geneesmiddel per patiënt: €19.000

Jaarimpact voor de Nederlandse publieke medicijn budgetten: €7.700.000 (dekking voor alle patiënten)

De totale uitgaven voor geneesmiddelen zijn in Nederland €5,9 miljard per jaar.

Waardeer de consequenties voor het budget/ kosten van de interventie:

Kies een getal tussen 0 (hoge bijkomende uitgaven) en 100 (flinke besparingen).

Consequenties voor het budget/ kosten van de interventie



Extra kosten per extra centimeter in de uiteindelijke lengte: € 16.000.

Extra kosten per gewonnen QALY: € 165.000.

Waardeer de kosten-effectiviteit van de interventie:

Kies een getal tussen 0 (niet kosten effectief) en 100 (zeer kosten effectief).

Kosteneffectiviteit van de interventie



Extra kosten per patiënt: € 794

Deze kosten zijn niet gemaakt door de patiënt, maar dit zijn kosten die gemaakt moeten worden voor de behandeling van de patiënt, bijv. training door de verpleegkundige, poliklinische bezoeken en röntgenfoto's gedurende 6 jaar (exclusief kosten van geneesmiddelen).

Waardeer de impact van de interventie op andere uitgaven:

Kies een getal tussen 0 (hoge bijkomende uitgaven) en 100 (flinke besparingen).

Impact op andere uitgaven



Kwaliteit van het bewijs:

Epidemiologie: beperkte statistische informatie;

Klinische gegevens: beperkte rapportage van bijwerkingen;

PRO (patiënt gerapporteerde uitkomsten): onvolledige rapportage van de vragenlijst aspecten;

Economische impact: sommige kenmerken van het model zijn onduidelijk;

Budgettaire gevolgen: geen sensitiviteitsanalyse gemeld;

Waardeer de compleetheit en eenduidigheid van het gerapporteerde bewijs:

Kies een getal tussen 0 (veel missende informatie/ niet eenduidig) en 100 (compleet en eenduidig).

Het gerapporteerde bewijs



Epidemiologie: 4 placebo gecontroleerde RCTs (N= 42-104, 1 in Canada, 3 in de VS); Observationeel gecontroleerde studies (N= 26-123, 1 in Duitsland, 1 in Griekenland, 1 in Israël, 3 in Italië);

Klinische gegevens: onzekerheid over de uiteindelijke lengte winst, hoog verloop in belangrijke RCT;

PRO: tussentijdse analyse van een subgroup van de deelnemers aan een niet anonieme RCT;

Economische evaluatie: twijfelachtige uitkomst-kosten per cm van de uiteindelijke lengte, geen bijwerkingen kosten, zwakke bruikbaarheidsgegevens;

Invloed budget: model neemt aan dat alle meisjes in Canada worden behandeld.

Waardeer de relevantie en betrouwbaarheid van deze gegevens:

Kies een getal tussen 0 (lage relevantie/ geldigheid) en 100 (hoge relevantie/ geldigheid).

Het bewijs



EIND VAN DE VRAGENLIJST

In dit onderzoek zijn drie verschillende antwoordschalen gebruikt. Hiervan heeft u er twee te zien gekregen. Heeft u een voorkeur voor een van de twee gebruikte antwoordschalen?

Choose one of the following answers

- ☐ 0-3 (4 punts schaal)
- ☐ 0-6 (7 punts schaal)
- ☐ 0-100 (visuele schaal)

- ☐ Geen voorkeur

Heeft u nog opmerkingen?

Answer



Bedankt voor het invullen van de vragenlijst! Voor vragen kunt u contact opnemen met f.c.uelderink@student.utwente.nl

Appendix 5: EVIDEM framework Version 3 (Scale 0-6)

Voor u ligt een vragenlijst over een interventie voor groeihormonen bij jonge meisjes met het syndroom van Turner. Het betreft hier een hypothetische situatie.

Deze vragenlijst is bedoeld voor studenten aan de Universiteit Twente met de studies Gezondheidswetenschappen, Technische Geneeskunde of Biomedische Technologie, en studenten van de studie Verpleegkunde aan het Saxion Enschede.

In de eerste helft van de vragenlijst wordt er gevraagd hoe belangrijk u elk criterium vindt in de waardering van een gezondheidsinterventie. In het tweede deel wordt u gevraagd om aan te geven hoe u de ziekte van Turner waardeert op deze criteria.

In deze vragenlijst zullen alle vragen 2 keer voorkomen, maar met een verschillende antwoordschaal.

Deze vragenlijst bestaat uit 44 vragen en zal ongeveer een kwartier de tijd kosten.

De hypothetische situatie in dit onderzoek is het toedienen van groeihormonen (GH) bij jonge meisjes met het syndroom van Turner.

Medicijn: polypeptide hormoon

Indicatie: behandeling van een korte gestalte bij meisjes met het syndroom van Turner

Behandeling: subcutane injectie 3 tot 7 dagen per week

Duur van de interventie: Moet worden vastgesteld. Start direct als groeistoornis is geconstateerd tot een bevredigende lengte bereikt is (6 jaar behandeling vanaf 10 jaar)

Vergelijking(en): Zonder behandeling

Economische last van de ziekte: Geen gegevens beschikbaar.

Wat is uw leeftijd?

Answer

Only numbers may be entered in this field.

Wat studeert u?

Choose one of the following answers

- ☐ Technische Geneeskunde
- ☐ Biomedische Technologie
- ☐ Gezondheidswetenschappen
- ☐ Verpleegkunde

Vult u hier uw email adres of naam in, zodat we de resultaten van uw 2 ingevulde vragenlijsten aan elkaar kunnen koppelen.

Answer

WEGING VAN DE CRITERIA

Geef aan hoe belangrijk je elk criterium vind in de beoordeling van een gezondheidsinterventie. Als je gelooft dat een criterium systematisch niet zou moeten worden beschouwd, ken een 0 toe.

Ernst van de ziekte

	1 (niet belangrijk)	2	3	4	5 (belangrijk)	0 (moet systematisch niet worden beschouwd)
Ernst van de ziekte	☐	☐	☐	☐	☐	☐

? De ernst van de aandoening van de patiënten behandeld met de voorgestelde interventie (of de ernst van de aandoening die voorkomen dient te worden) in termen van de sterfte, invaliditeit, impact op de kwaliteit van leven, het verloop van de ziekte (d.w.z. spoedeisend, verloop klinische fases).

Grootte van de beïnvloede populatie

	1 (niet belangrijk)	2	3	4	5 (belangrijk)	0 (moet systematisch niet worden beschouwd)
Grootte van de beïnvloede populatie:	☐	☐	☐	☐	☐	☐

? Aantal mensen getroffen door de aandoening (behandeld of voorkomen door de voorgestelde interventie) in een bepaalde populatie op een bepaald tijdstip; dit kan worden uitgedrukt als jaarlijks aantal nieuwe gevallen (jaarlijkse incidentie) en/ of deel van de getroffen bevolking op een bepaald punt in de tijd (prevalentie).

Aansluiting tussen de interventie en de klinische praktijk richtlijnen

	1 (niet belangrijk)	2	3	4	5 (belangrijk)	0 (moet systematisch niet worden beschouwd)
Klinische praktijk richtlijnen:	☐	☐	☐	☐	☐	☐

? Overeenkomst tussen de voorgestelde interventie (of van soortgelijke alternatieven) met de huidige consensus van deskundigen over wat bijdraagt aan de **state-of-the-art** voor de management van de beoogde aandoening; richtlijnen worden gewoonlijk ontwikkeld via een expliciet proces en zijn bedoeld om de klinische praktijk te verbeteren.

Beperkingen van vergelijkbare interventies (onvervulde behoeften)

	1 (niet belangrijk)	2	3	4	5 (belangrijk)	0 (moet systematisch niet worden beschouwd)
Beperkingen van vergelijkbare interventies (onvervulde behoeften)	☐	☐	☐	☐	☐	☐

? Tekortkomingen van vergelijkbare interventies in hun vermogen om de aandoening te voorkomen, genezen of de gevolgen te verminderen; ook tekortkomingen met betrekking tot veiligheid, patiënt-gerapporteerde uitkomsten en geschiktheid.

Verbetering van de werkzaamheid/ effectiviteit

	1 (niet belangrijk)	2	3	4	5 (belangrijk)	0 (moet systematisch niet worden beschouwd)
Verbetering van de werkzaamheid/ effectiviteit	☐	☐	☐	☐	☐	☐

? De mogelijkheden van de voorgestelde interventie om te komen tot een gewenste (positieve) verandering in de tekenen, symptomen en het verloop van de beoogde aandoening in aanvulling op de gunstige veranderingen veroorzaakt door andere, bestaande interventies.

Inclusief gegevens over werkzaamheid en de effectiviteit , indien beschikbaar.

Verbetering van de veiligheid

	1 (niet belangrijk)	2	3	4	5 (belangrijk)	0 (moet systematisch niet worden beschouwd)
Verbetering van de veiligheid	☐	☐	☐	☐	☐	☐

? De mogelijkheden van de voorgestelde interventie om tot een vermindering van de aan de interventie gerelateerde schadelijke of ongewenste effecten op de aandoening te komen, in vergelijking met andere interventies.

Verbetering van de patiënt-ervaren gezondheid/ patiënt gerapporteerde uitkomsten

	1 (niet belangrijk)	2	3	4	5 (belangrijk)	0 (moet systematisch niet worden beschouwd)
Verbetering van de patiënt-ervaren gezondheid/ patiënt gerapporteerde uitkomsten:	☐	☐	☐	☐	☐	☐

? Capaciteit van de voorgestelde interventie om te komen tot positieve veranderingen in door de patiënt gerapporteerde uitkomsten (bijv. kwaliteit van leven) in aanvulling op gunstige veranderingen veroorzaakt door andere interventies; inclusief verbetering in het comfort van patiënten.

Aard van de medische dienst (therapeutisch)

	1 (niet belangrijk)	2	3	4	5 (belangrijk)	0 (moet systematisch niet worden beschouwd)
Aard van de medische dienst (therapeutisch):	☐	☐	☐	☐	☐	☐

? Aard van het klinische voordeel veroorzaakt door de voorgestelde interventie op patiënt-niveau (bijv. verlichting van de symptomen, het verlengen van de levensduur, genezing).

Consequenties voor het budget/ kosten van de interventie

	1 (niet belangrijk)	2	3	4	5 (belangrijk)	0 (moet systematisch niet worden beschouwd)
Consequenties voor het budget/ kosten van de interventie:	☐	☐	☐	☐	☐	☐

? Netto effect van de vergoeding van de interventie op het budget van de basisverzekering (exclusief overige uitgaven). Dit is het verschil tussen de verwachte uitgaven voor de voorgestelde interventie en de potentiële kostenbesparingen die kunnen voortvloeien uit de vervanging van andere interventie(s) die nu vallen onder de basisverzekering. Beperkt tot de kosten van de interventie (bijv. aanschafkosten, implementatie- en onderhoudskosten).

Kosten-effectiviteit van de interventie

	1 (niet belangrijk)	2	3	4	5 (belangrijk)	0 (moet systematisch niet worden beschouwd)
Kosten-effectiviteit van de interventie:	☐	☐	☐	☐	☐	☐

? Verhouding van de extra kosten van de voorgestelde interventie tot het bijkomende voordeel in vergelijking met alternatieven. Voordeel kan worden uitgedrukt als het aantal gebeurtenissen die zijn vermeden, gewonnen levensjaren, verbeterde kwaliteit van de gewonnen levensjaren, extra pijn-vrije dagen, etc.

Impact op andere uitgaven (bijv. ziekenhuisopname, invaliditeit)

	1 (niet belangrijk)	2	3	4	5 (belangrijk)	0 (moet systematisch niet worden beschouwd)
Impact op andere uitgaven:	☐	☐	☐	☐	☐	☐
? Impact van het verstrekken van dekking voor de voorgestelde interventie op andere uitgaven (uitgezonderd interventie kosten), zoals ziekenhuisopname, consulten bij specialisten, bijwerkingen, langdurige zorg, kosten invaliditeit, verlies van productiviteit, tijd van de verzorger, etc. Volledigheid en samenhang van het gerapporteerde bewijs (het voldoen aan de wetenschappelijke rapportage normen en eenduidigheid met de bronnen)						
	1 (niet belangrijk)	2	3	4	5 (belangrijk)	0 (moet systematisch niet worden beschouwd)
Volledigheid en samenhang van het gerapporteerde bewijs:	☐	☐	☐	☐	☐	☐
? Mate waarin de gerapporteerde gegevens over de voorgestelde interventie compleet zijn (dat wil zeggen, het voldoen aan wetenschappelijke normen voor rapportage) en in overeenstemming zijn met de aangehaalde bronnen. Relevantie en geldigheid van het bewijs (relevant voor de besluitnemers & voldoen aan de wetenschappelijke normen)						
	1 (niet belangrijk)	2	3	4	5 (belangrijk)	0 (moet systematisch niet worden beschouwd)
Relevantie en geldigheid van het bewijs:	☐	☐	☐	☐	☐	☐
? Mate waarin het bewijs voor de voorgestelde interventie relevant is voor het besluitvormend orgaan (in termen van populatie, ziekte stadium, vergelijkbare interventies, resultaten, etc.) en in overeenstemming met de wetenschappelijke normen (d.w.z. onderzoeksopzet, etc.) en conclusies (overeenstemming van resultaten tussen de studies). Dit is inclusief het in aanmerking nemen van onzekere factoren (bijvoorbeeld tegenstrijdige resultaten tussen de onderzoeken, beperkt aantal onderzoeken en patiënten).						

Let op: Dezelfde 14 vragen als hiervoor zullen nu nogmaals volgen, maar met een andere antwoordschaal.

De ziekte van Turner is een vrouw-specifieke genetische aandoening die wordt gekenmerkt door een korte gestalte, cardiovasculaire gebreken, het ontbreken van de puberteit, onvruchtbaarheid, verhoogd risico op diabetes, defecten in visueel-ruimtelijke organisatie en non-verbale probleemoplossing, psychosociale problemen en een verminderde levensverwachting.

Waardeer de ernst van deze ziekte:

	0 (niet ernstig)	1	2	3	4	5	6 (heel ernstig)
Ernst van de ziekte	☐	☐	☐	☐	☐	☐	☐

De prevalentie is 40/100.000 vrouwelijke volwassenen.

Waardeer de grootte van de getroffen populatie:

	0 (hele zeldzame ziekte)	1	2	3	4	5	6 (veel voorkomende ziekte)
Grootte van de getroffen populatie	☐	☐	☐	☐	☐	☐	☐

de internationale richtlijnen voor de ziekte van Turner schrijven voor dat de groeihormoon behandeling overwogen moet worden zodra de groeistoornis wordt geconstateerd en potentiële risico's/ voordelen zijn besproken met de patiënt / familie. Behandelen totdat een bevredigende lengte is bereikt.

Waardeer de aansluiting van de interventie op de bestaande klinische praktijkrichtlijnen:

	0 (niet aanbevolen)	1	2	3	4	5	6 (sterke aanbeveling)
Klinische praktijkrichtlijnen:	☐	☐	☐	☐	☐	☐	☐

Er is geen andere therapeutische interventie geïndiceerd om korte gestalte te behandelen bij het syndroom van Turner.

Waardeer de beperkingen van vergelijkbare interventies (onvervulde behoeften):

	0 (geen of heel weinig beperkingen - onvervulde behoeften)	1	2	3	4	5	6 (belangrijke beperkingen - onvervulde behoeften)
Vergelijkbare interventies:	☐	☐	☐	☐	☐	☐	☐

Uitkomsten van de interventie:

In 4 placebo controlled RCTs was het **verschil met onbehandelde patiënten 7 cm.**

In observationeel gecontroleerde studies was het **verschil met de controlegroep = 2,1-6,8 cm.**

Waardeer de verbetering van de werkzaamheid/ effectiviteit:

	-6 (lager dan zonder behandeling)	-5	-4	-3	-2	-1	0 (geen verbetering)	1	2	3	4	5	6 (grote verbetering t.o.v. geen behandeling)
Werkzaamheid/ effectiviteit:	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

De meest voorkomende bijwerkingen van het medicijn zijn: Chirurgie (50%), oorproblemen (6% tot 47%) gewrichts- (13,5%) en respiratoire (11%) aandoeningen, sinusitis (18,9%).

Waardeer de veiligheid van het medicijn:

	-6 (veel minder veilig dan zonder behandeling)	-5	-4	-3	-2	-1	0 (geen verbetering)	1	2	3	4	5	6 (veel veiliger dan geen behandeling)
Veiligheid:	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Er zijn geen significantie verschillen in de verbetering van de ervaren gezondheid.

Waardeer de verbetering van de ervaren gezondheid/ patiënt gerapporteerde uitkomsten:

	-6 (lager dan zonder behandeling)	-5	-4	-3	-2	-1	0 (geen verbetering)	1	2	3	4	5	6 (grote verbetering t.o.v. geen behandeling)
De ervaren gezondheid/ patiënt gerapporteerde uitkomsten:	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Type voordeel:

Het doel van de behandeling is het stimuleren van de groei en het verbeteren van het psychosociale welzijn (lengte winst 7 cm, patiënt gerapporteerde uitkomsten gegevens beperkt en niet eenduidig).

Waardeer de aard van de medische dienst (therapeutisch):

Epidemiologie: 4 placebo gecontroleerde RCTs (N= 42-104, 1 in Canada, 3 in de VS); Observationeel gecontroleerde studies (N= 26-123, 1 in Duitsland, 1 in Griekenland, 1 in Israël, 3 in Italië);

Klinische gegevens: onzekerheid over de uiteindelijke lengte winst, hoog verloop in belangrijke RCT;

PRO: tussentijdse analyse van een subgroup van de deelnemers aan een niet anonieme RCT;

Economische evaluatie: twijfelachtige uitkomst-kosten per cm van de uiteindelijke lengte, geen bijwerkingen kosten, zwakke bruikbaarheidsgegevens;

Invloed budget: model neemt aan dat alle meisjes in Canada worden behandeld.

Waardeer de relevantie en betrouwbaarheid van deze gegevens:

	6 (hoge relevantie/ geldigheid)	5	4	3	2	1	0 (lage relevantie/ geldigheid)
Relevantie en geldigheid van het bewijs:	☐	☐	☐	☐	☐	☐	☐

EIND VAN DE VRAGENLIJST

In dit onderzoek zijn drie verschillende antwoordschalen gebruikt. Hiervan heeft u er twee te zien gekregen. Heeft u een voorkeur voor een van de twee gebruikte antwoordschalen?

Choose one of the following answers

- ☐ 0-3 (4 punts schaal)
- ☐ 0-6 (7 punts schaal)
- ☐ 0-100 (visuele schaal)
- ☐ Geen voorkeur

Heeft u nog opmerkingen?

Answer

Bedankt voor het invullen van de vragenlijst! Voor vragen kunt u contact opnemen met f.c.uelderink@student.utwente.nl