Review of proposed outcome measures for cNF: tumor size, PROs, and global assessment of change (GAC)

Scott Plotkin, MD, PhD
Dominique Pichard, MD

REiNS 2019 Winter Meeting
December 9, 2019

Validating measurement techniques for cutaneous neurofibromas in clinical trials

Raquel D. Thalheimer, Vanessa L. Merker, Ina Ly, Sharad K. Verma, Amanda H. Champlain, Jennifer L. Sawaya, Naomi Askenazi, Justin T. Jordan, Alona Muzikansky, Fernanda Sakamoto, Jaishri O. Blakeley, R. Rox Anderson, Scott R. Plotkin

REiNS 2019 Winter Meeting
December 9, 2019

Disclosures and Funding

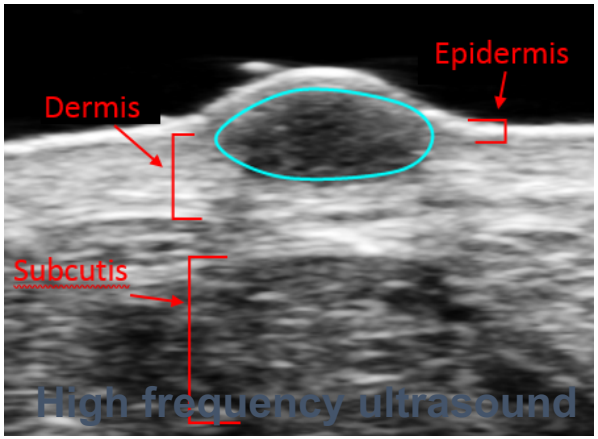
- Co-founder of NFlection Therapeutics.
- Consulting with AstraZeneca
- Study is funded by the Neurofibromatosis Therapeutic Acceleration Program (NTAP) at Johns Hopkins University.

Validating assessments of cNF size

- A major limitation to evaluating either novel interventions or drugs is the ***inability to assess cNF size*** with a reliable and reproducible measurement tool.
- Prospective, single center study.
- Eligibility: NF1 diagnosis, ≥ 18 years old, at least 6 visible cNFs, ability to read English, able to tolerate imaging
- Tumor measurements were acquired using digital calipers, 3D photography, and high frequency ultrasound (HFUS) in a single session.

Digital calipers

Canfield Scientific, Vectra H1



Validating assessments of cNF size

- Aims:
  - Determine the intra-rater and inter-rater reliability of HFUS measurements of cNF volume at baseline
  - Determine the accuracy of HFUS measurements by comparing them to digital caliper measurements and 3D photography
- Statistical analysis
  - Linear and volumetric assessments were compared using **intraclass correlation coefficient (ICC)** to determine the intra- and inter-rater reliability of each technique.

Demographics of reliability set

Feature	Value
Participants (n)	10
Age (median, years)	53 (range 36-67)
Sex (female)	5 (50%)
Ethnicity Non-Hispanic Hispanic	 8 (80%) 1 (10%) One response omitted
Race White Asian	 9 (90%) 1 (10%)
Cutaneous neurofibromas (n)	57
Median diameter (range) – mm < 5 mm ≥ 5 mm	4.2 (2.4-11.4) 33 (58%) 24 (42%)

Reliability Assessment

- Baseline visit details (N=10)
  - 3 researchers acquiring images (**acquisition**) – inter-rater
    - 1 researcher acquiring 3 images per tumor – intra-rater
  - 3 researchers assessing images (**measurement**)
    - 1 researcher measuring 3 times per tumor
  - 60 scans per patient x 10 patients = 600 scans
  - 300 measurements per patient x 10 patients = 3000 measurements

Reliability of HFUS, 3D camera, and calipers

Intraclass Correlation Coefficient (ICC)

	Image Acquisition Reliability		Image Analysis Reliability	
	Intra-rater ICC	**Inter-rater ICC**	**Intra-rater ICC**	**Inter-rater ICC**
HFUS				
Volume	0.98	0.97	0.99	0.98
Width	0.98	0.98	0.98	0.99
Depth	0.98	0.97	0.99	0.97
3D Camera				
Volume	0.97	0.95	0.98	0.98
Width Manual	0.97	0.96	0.97	0.98
Width Script	0.96	0.96	0.98	0.97
Length Manual	0.97	0.96	0.98	0.98
Length Script	0.97	0.87	0.91	0.93
Surface Area	0.98	0.97	0.99	0.98
Calipers				
Volume			0.90	0.77
Width			0.96	0.88
Length			0.93	0.85
Height			0.81	0.62

ICC measures reproducibility

Shrout-Fleiss reliability: random set

- 57 Tumors total:
  - 33 small tumors (<5mm)
  - 24 large tumors (≥5mm)

ICC	Reliability
<.5	Poor
0.5-.75	Moderate
0.75-0.9	Good
0.9-1.0	Excellent

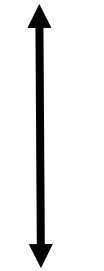
Reliability Observations

- ICC “excellent” for all 3D photography and HFUS observations
- ICC least strong for inter-rater caliper measurements (particularly height), however still in “moderate” range

Determining the minimal detectable change

- The **coefficient of variation** (CV) is defined as the ratio of the standard deviation to the mean: It shows the extent of variability in relation to the mean of the population.
- Study Participants 1-10
- 57 Tumors total:
  - 33 small tumors (<5mm)
  - 24 large tumors (≥5mm)

COV
≥35%
30-34%
26-29%
21-25%
16- 20%
11-15%
≤ 11%



	HFUS COV (mean)			3D Camera COV (mean)						Calipers COV (mean)			
	Volume	Width	Depth	Volume	Width	Length	Max width (script)	Length (script)	Surface area	Volume	Length	Width	Height
Small	35.7%	13.4%	11.6%	57.2%	11.8%	10.7%	10.2%	10.9%	18.7%	63.2%	17.8%	19.0%	38.7 %
Large	17.8%	5.9%	5.2%	18.6%	7.0%	6.6%	7.6%	9.9%	12.2%	33.4%	13.0%	10.6%	25.0 %
All	29.5%	10.4%	9.2%	43.2%	9.7%	8.9%	9.1%	10.3%	16.3%	51.8%	15.8%	15.8%	33.0 %

COV and Suggested Threshold for Imaging Response Overall

	Small COV	Large COV	All COV	Proposed threshold for clinical response
Calipers				
Width	19.0%	10.6%	15.8%	30%
Length	17.8%	13.0%	15.8%	30%
Height	38.7%	25.0%	33.0%	65%
Volume	63.2%	33.4%	51.8%	100%
3D Camera				
Width (manual)	11.8%	7.0%	9.7%	20%
Width (script)	10.2%	7.6%	9.1%	18%
Length (manual)	10.7%	6.6%	8.9%	18%
Length (script)	10.9%	9.9%	10.3%	20%
Surface Area	18.7%	12.2%	16.3%	33%
Volume	57.2%	18.6%	43.2%	85%
HFUS				
Width	13.4%	5.9%	10.4%	20%
Depth	11.6%	5.2%	9.2%	20%
Volume	35.7%	17.8%	29.5%	60%

Threshold
≥35%
30-34%
26-30%
21-25%
16- 20%
11-15%
≤ 11%

Comparing measurement tools

	Cost	Reliability	Best use	Threshold
Digital Calipers	~ $100	Good	Few tumors Large tumors Limited budget	30% (linear)
3D Camera	~$14K	Excellent	Multiple tumors assessing color Local or systemic rx	20% (linear) 33% (area)
HFUS	> $225K	Excellent	Few tumors Small tumors Prevention trials Visualize beneath skin	20% (linear)

Next steps

- Establish criteria for measurement of cutaneous neurofibromas in clinical trials
- Longitudinal assessment of cNF to determine natural history of tumor growth/shrinkage
- Apply measurement criteria in a treatment trial (e.g., selumetinib for cNF, NCT02839720, PI: Bruce Korf)

Patient Reported Outcomes

Dominique Pichard, MD; Pam Wolters, PhD; Dawn Siegel, MD, PhD; Khaled Ezzedine, MD; Betty Schorry, MD; and Carolina Barnett Tapia, MD

Massachusetts General Hospital

Scott Plotkin, MD, PhD
Vanessa L. Merker, PhD
Ina Ly, MD
Amanda H. Champlain, MD
Jennifer L. Sawaya, MD
Alona Muzikansky, MA
Fernanda Sakamoto, MD, PhD
R. Rox Anderson, MD
Naomi Askenazi

Johns Hopkins University

Jaishri Blakeley, MD
Sharad Verma, PhD

NTAP
Neurofibromatosis
Therapeutic Acceleration Program
at Johns Hopkins

Response Evaluation in Neurofibromatosis & Schwannomatosis
REiNS

Process

Review the literature for existing PROs that are specific for assessing the skin

- Review of the literature: Skin specific PRO instruments identified
  - Skindex & teen-skindex
  - Dermatology life quality index (DLQI) & CDLQI
  - *Adjusted NF QOL (Hilda Crawford)*
  - *Itch scales*
  - *Pain scales*

2 at Os

ed Outcomes Rating Acceptance Tool for Endpoints
or REiNS Committee use only)

Disease-Specific QOL ☐Pain ☐Functional Disability ☐_____

_____

_____Date:_____

pporting its use in neurofibromatosis trials
rmation but needs more work
otential

0=No/poor data/information
*Half ratings (.5, 1.5, 2.5) can be used if needed

Rating Criteria	Rating (0-3): use in NF trials
1. Patient characteristics: Age range (e.g., child, adolescent, adult) Normative groups (e.g., general, NF, oncology, other, # subjects)	
2. Used in published studies: Number and types of studies (e.g., descriptive, clinical trials)	
3. Domains assessed/Item content: Number/description (e.g., physical, social, emotional, cognitive)	
4. Scores available: Item response format (e.g., Likert scale, visual analog scale) Types of scores (e.g., raw, standardized, domain, total)	
5. Psychometric Data: Reliability (e.g., internal consistency, test/retest) Validity (e.g., construct, discriminative) Factor analysis	
6. Feasibility: Cost Length (number of items) Ease of administration Recall period assessed (e.g., 1 week, 24 hours) Availability in different languages/International use	
Overall Impression for use in NF Clinical Trials (Pros/Cons):	Total (mean):

Level of Acceptance (Committee decision):
___ Primary outcome measure
___ Secondary outcome measure
___ Not acceptable at this time/further information needed (specify)
___ Not acceptable (no further review)

Committee notes/comments/additional information needed/plan:

SKINDEX

HOW OFTEN DURING THE PAST FOUR WEEKS DO THESE STATEMENTS DESCRIBE YOU?	NEVER	RARELY	SOMETIMES	OFTEN	ALL THE TIME
1. My skin hurts	☐1	☐2	☐3	☐4	☐5
2. My skin condition affects how well I sleep	☐1	☐2	☐3	☐4	☐5
3. I worry that my skin condition may be serious	☐1	☐2	☐3	☐4	☐5
4. My skin condition makes it hard to work or do hobbies	☐1	☐2	☐3	☐4	☐5
5. My skin condition affects my social life	☐1	☐2	☐3	☐4	☐5
6. My skin condition makes me feel depressed	☐1	☐2	☐3	☐4	☐5
7. My skin condition burns or stings	☐1	☐2	☐3	☐4	☐5
8. I tend to stay at home because of my skin condition	☐1	☐2	☐3	☐4	☐5
9. I worry about getting scars from my skin condition	☐1	☐2	☐3	☐4	☐5
10. My skin itches	☐1	☐2	☐3	☐4	☐5
11. My skin condition affects how close I can be with those I love	☐1	☐2	☐3	☐4	☐5
12. I am ashamed of my skin condition	☐1	☐2	☐3	☐4	☐5
13. I worry that my skin condition may get worse	☐1	☐2	☐3	☐4	☐5
14. I tend to do things by myself because of my skin condition	☐1	☐2	☐3	☐4	☐5
15. I am angry about my skin condition	☐1	☐2	☐3	☐4	☐5
16. Water bothers my skin condition (bathing, washing hands)	☐1	☐2	☐3	☐4	☐5
17. My skin condition makes showing affection difficult	☐1	☐2	☐3	☐4	☐5
18. I worry about side-effects from skin medications / treatments	☐1	☐2	☐3	☐4	☐5
19. My skin is irritated	☐1	☐2	☐3	☐4	☐5
20. My skin condition affects my interactions with others	☐1	☐2	☐3	☐4	☐5

S
K
I
N
D
E
X

HOW OFTEN DURING THE PAST 4 WEEK DO THESE STATEMENTS DESCRIBE YOU?	NEVER	RARELY	SOMETIMES
21. I am embarrassed by my skin condition	☐1	☐2	☐3
22. My skin condition is a problem for the people I love	☐1	☐2	☐3
23. I am frustrated by my skin condition	☐1	☐2	☐3
24. My skin is sensitive	☐1	☐2	☐3
25. My skin condition affects my desire to be with people	☐1	☐2	☐3
26. I am humiliated by my skin condition	☐1	☐2	☐3
27. My skin condition bleeds	☐1	☐2	☐3
28. I am annoyed by my skin condition	☐1	☐2	☐3
29. My skin condition interferes with my sex life	☐1	☐2	☐3
30. My skin condition makes me tired	☐1	☐2	☐3

Skindex PRO RATE

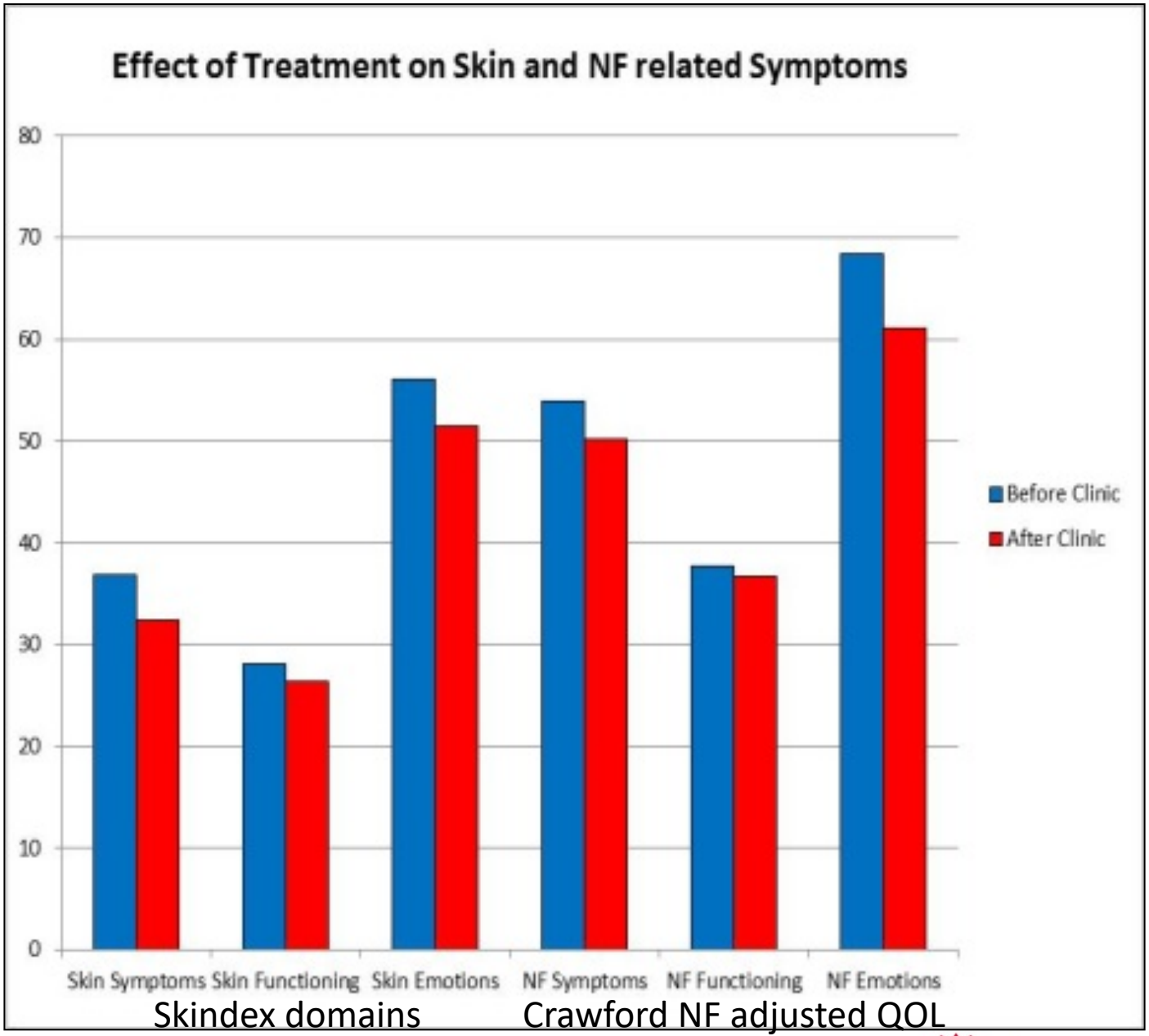
Overall impression: 2.54/3.0

PROS

- Ages 12-17, ≥18
- Used widely in general derm.
- Qs appropriate for cNFs
- Feasibility

CONS

- No interventional trial data
- Questions not relevant to NF1
- “my skin” or “my skin condition”

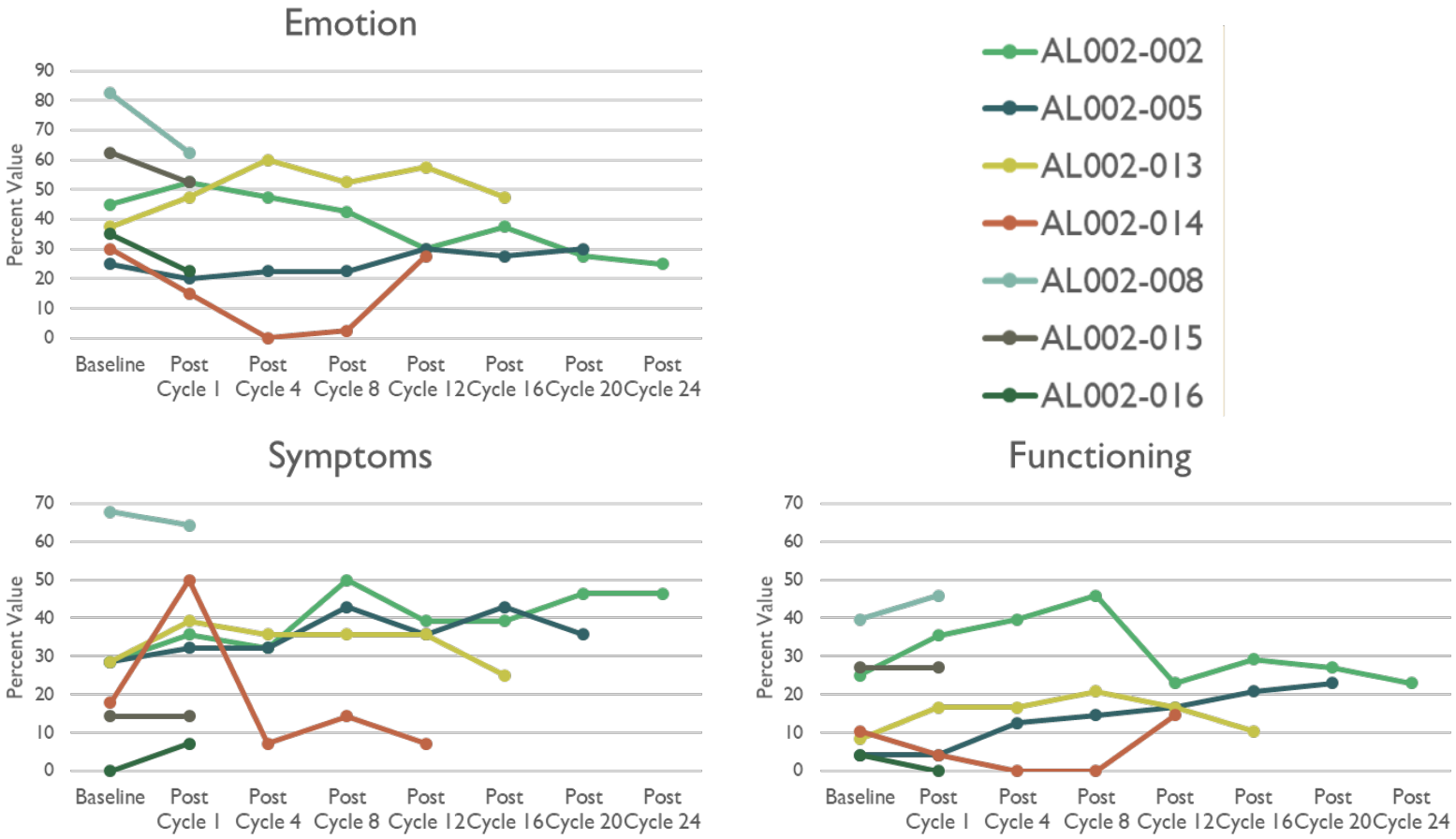


Skindex domains

Crawford NF adjusted QOL domains

Health
Northern Sydney
Local Health District

Skindex Scores



DLQI

1.	Over the last week, how **itchy**, **sore**, **painful** or **stinging** has your skin been?	Very much A lot A little Not at all	❒ ❒ ❒ ❒		
2.	Over the last week, how **embarrassed** or **self conscious** have you been because of your skin?	Very much A lot A little Not at all	❒ ❒ ❒ ❒		
3.	Over the last week, how much has your skin interfered with you going **shopping** or looking after your **home** or **garden**?	Very much A lot A little Not at all	❒ ❒ ❒ ❒	Not relevant	❒
4.	Over the last week, how much has your skin influenced the **clothes** you wear?	Very much A lot A little Not at all	❒ ❒ ❒ ❒	Not relevant	❒
5.	Over the last week, how much has your skin affected any **social** or **leisure** activities?	Very much A lot A little Not at all	❒ ❒ ❒ ❒	Not relevant	❒
6.	Over the last week, how much has your skin made it difficult for you to do any **sport**?	Very much A lot A little Not at all	❒ ❒ ❒ ❒	Not relevant	❒
7.	Over the last week, has your skin prevented you from **working** or **studying**?	Yes No	❒ ❒	Not relevant	❒
	If "No", over the last week how much has your skin been a problem at **work** or **studying**?	A lot A little Not at all	❒ ❒ ❒		
8.	Over the last week, how much has your skin created problems with your **partner** or any of your **close friends** or **relatives**?	Very much A lot A little Not at all	❒ ❒ ❒ ❒	Not relevant	❒
9.	Over the last week, how much has your skin caused any **sexual difficulties**?	Very much A lot A little Not at all	❒ ❒ ❒ ❒	Not relevant	❒
10.	Over the last week, how much of a problem has the **treatment** for your skin been, for example by making your home messy, or by taking up time?	Very much A lot A little	❒ ❒ ❒		

C
D
L
Q
I

Trouble with Skin

The aim of the questionnaire is to measure how much your skin problem has affected you OVER THE LAST WEEK. Please tick ✓ one box for each question.

OVER THE LAST WEEK

Very much ☐
Quite a lot ☐
A little ☐
Not at all ☐

How itchy, '**scratchy**', **sore** or **painful** has your skin been ?

OVER THE LAST WEEK

Very much ☐
Quite a lot ☐
A little ☐
Not at all ☐

How **upset** or **embarrassed**, **self conscious** or **sad** have you been because of your skin?

Very much ☐
Quite a lot ☐
A little ☐
Not at all ☐

How much has your skin affected your **friendships**?

Very much ☐
Quite a lot ☐
A little ☐
Not at all ☐

How much have you changed or worn **different** or **special clothes/ shoes** because of your skin?

Very much ☐
Quite a lot ☐
A little ☐
Not at all ☐

How much has your skin trouble affected **going out**, **playing** or **doing hobbies**?

Very much ☐
Quite a lot ☐
A little ☐
Not at all ☐

How much have you avoided **swimming** or **other sports** because of your skin trouble?

(C)DLQI PRO RATE

Overall impression: 2.5/3.0 (2.65/3)

PROS

- Ages 4-16, ≥16
- Used widely in derm. (>1000 pub)
- Qs appropriate for cNFs
- Feasibility

CONS

- No interventional trial data
- Multiple domains in single question
- Raw score, interpretation is not validated

Recommendations:

- Use a modified Skindex that changes the language of “my skin condition” to “my cutaneous neurofibromas”
  - Track sensitivity to change in an interventional study to determine if useful
  - If so, proceed with modified skindex, but if not modify the DLQI/CDLQI
- Add a Global Impression of Change

Global Impression of Change Scale - cNF

1. Think about your cutaneous neurofibromas now. Compare them to before you started taking the medicine for this study. Do you think your cutaneous neurofibromas are:

- ☐ 1 Very Much Improved
- ☐ 2 Much Improved
- ☐ 3 Minimally Improved
- ☐ 4 No Change
- ☐ 5 Minimally Worse
- ☐ 6 Much Worse
- ☐ 7 Very Much Worse

(Please check only one box)

2. Please describe any changes you have noticed:

Recommendations:

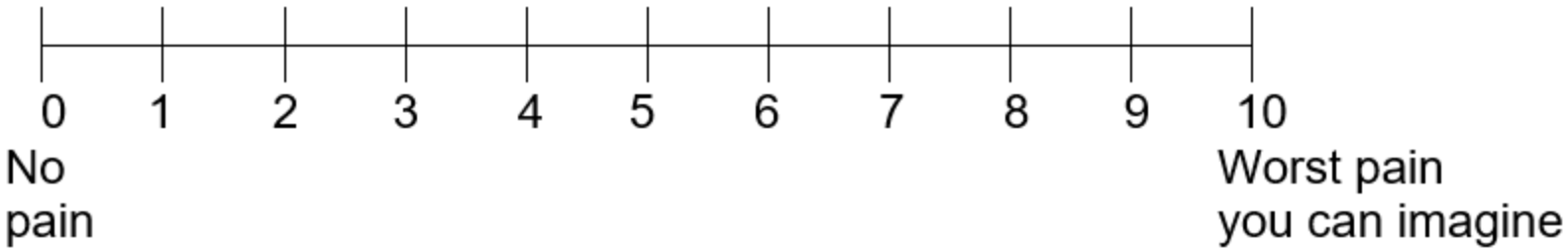
- Use a modified Skindex that changes the language of “my skin condition” to “my cutaneous neurofibromas”
  - Track sensitivity to change in an interventional study to determine if useful
  - If so, proceed with modified skindex, but if not modify the DLQI/CDLQI
- Add a Global Impression of Change
- Consider including NRS-11 (pain scale) for systemic trials

NRS-11

Numeric Rating Scale (NRS-11)

Below is a line with numbers from 0 to 10 where 0 means no pain and 10 means the worst pain you can imagine.

Please circle the one number that best describes your pain at its **worst** during the past week.



Acknowledgements:

- Pam Wolters, Dawn Siegel, Khaled Ezzedine, Betty Schorry, and Carolina Barnett Tapia from the REiNS PRO and cNF working groups

- Ashley Cannon, UAB

- All of the patients who have contributed to our knowledge and understanding through clinical trial participation

Towards a global assessment of change for cNF

December 9, 2019

Scott Plotkin, MD, PhD
Ashley Cannon, PhD, CGC
Dominique Pichard, MD
Claas Rohl

Response Evaluation In Neurofibromatosis Schwannomatosis
INTERNATIONAL COLLABORATION

Botox label

- The co-primary efficacy endpoints were the investigator's rating of glabellar line severity at maximum frown and the subject's global assessment of change in appearance of glabellar lines, both at Day 30 post-injection.
  - For the investigator rating, using a 4-point grading scale (0=none, 3=severe) a responder was defined as having a severity grade of 0 or 1.
  - For the subject's global assessment of change, the ratings were from +4 (complete improvement) to -4 (very marked worsening).

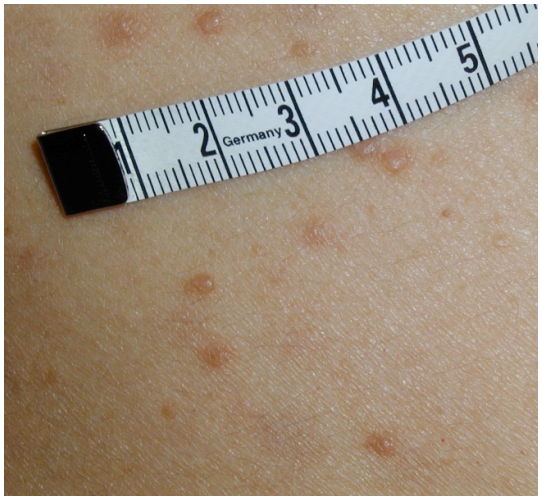
Table 5: Subject's Assessment of Change in Appearance of Glabellar Lines – Responder Rates (% and Number of Subjects with at Least Moderate Improvement)

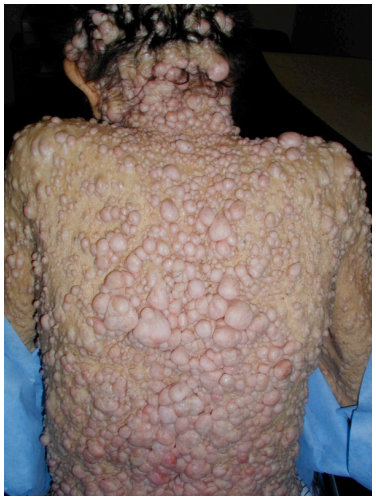
Day	BOTOX Cosmetic	Placebo	Difference[a]
7	82% 334/405	9% 12/132	73% (68, 80)
30[b]	89% 362/405	7% 9/132	83% (77, 88)
60	82% 330/403	4% 5/130	78% (73, 83)
90	63% 254/403	3% 4/128	60% (54, 66)
120	39% 157/403	1% 1/128	38% (33, 43)

[a] 95% confidence intervals are shown in parenthesis
[b] Day 30: Co-Primary Efficacy Time point, $p<0.001$

Creating a global assessment of change scale for cNF

- Decision NOT to use a severity scale given wide range of tumor burden (differing by 3 orders of magnitude)
- Recognizes that patients find benefit in less than 100% clearance (REiNS patient survey)
- Need to specify whether change is regional (e.g., due to local treatments) or whole body (e.g., due to systemic treatments)
- Attempt to correlate change in tumor size with assessment of GAC
- Validation is critical





Proposed global assessment of change scale for cNF size

For local treatment: “Compared to baseline, please rate global change in size within the field of treatment”

For systemic treatment: “Compared to baseline, please rate global change in size within the field of treatment”

- +3: significant improvement
- +2: moderate improvement

} Clinical response

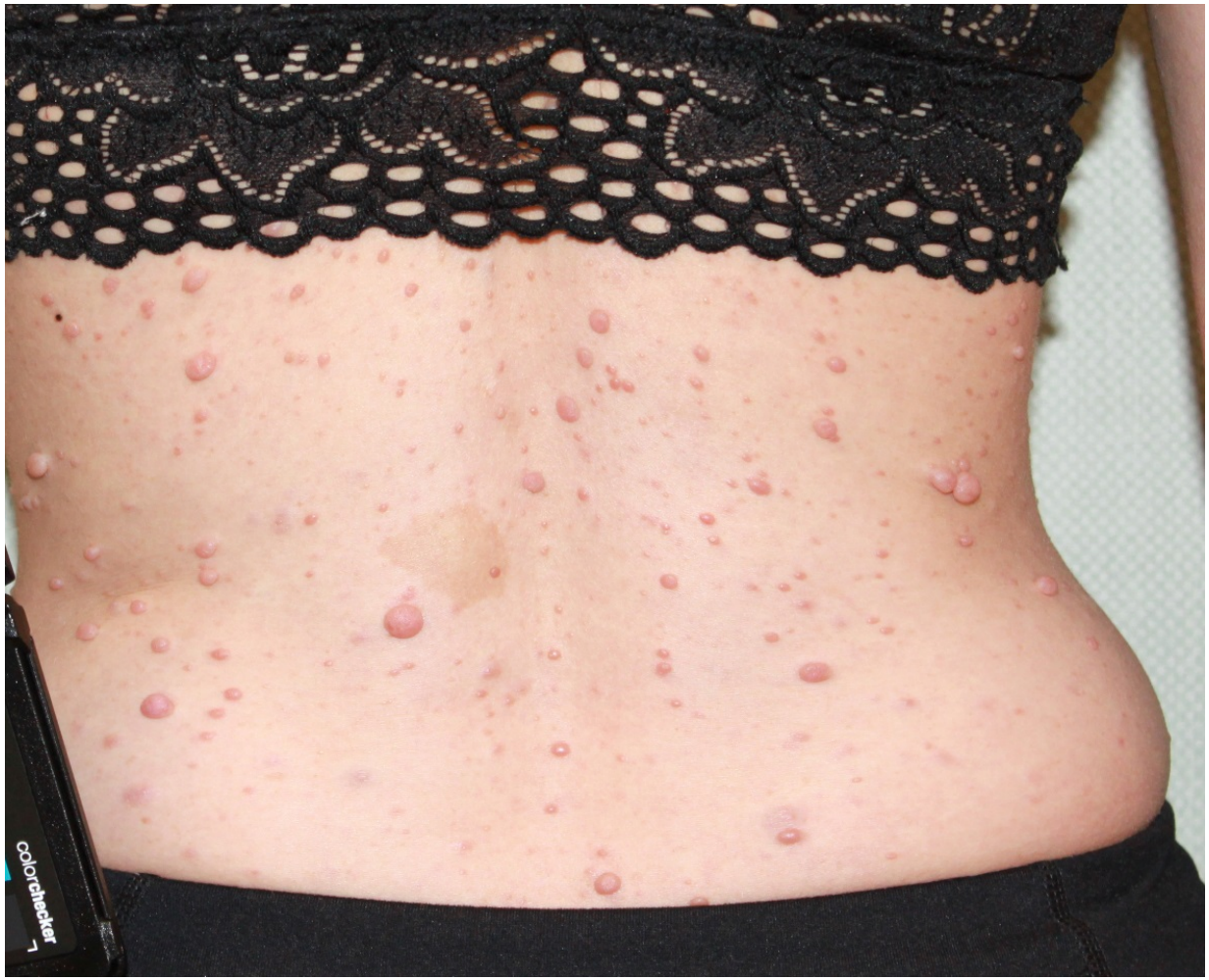
- +1: minimal improvement
- 0: no change
- -1: minimal worsening

} Stable

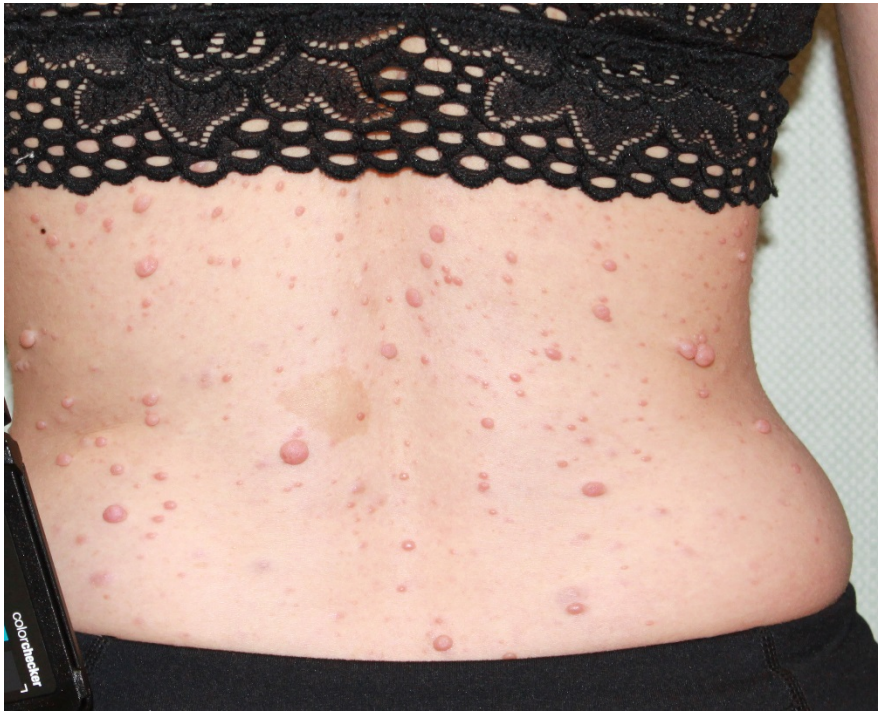
- -2: moderate worsening
- -3: significant worsening

} Clinical worsening

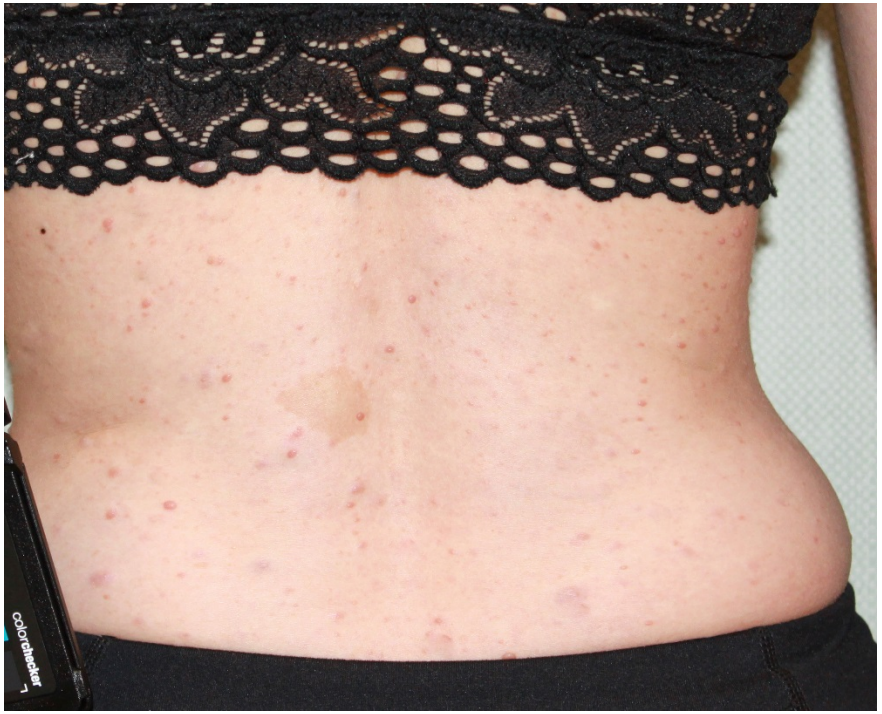
Index subject



S1. Simulation of whole body change



Baseline



After treatment

- +3: significant improvement
- +2: moderate improvement
- +1: minimal improvement

Clinical response

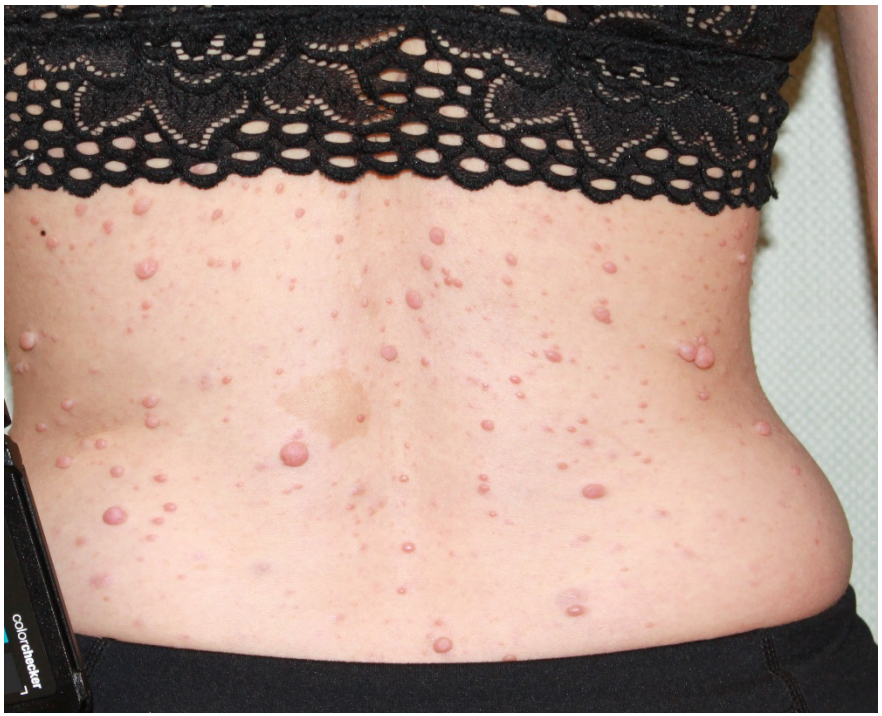
- 0: no change

Stable

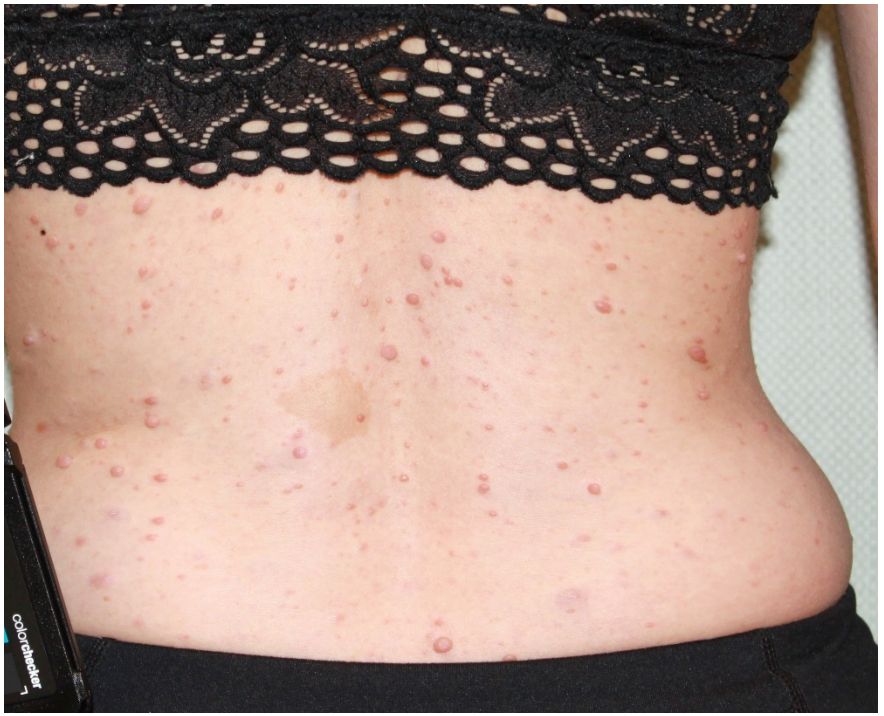
- -1: minimal worsening
- -2: moderate worsening
- -3: significant worsening

Clinical worsening

S2. Simulation of whole body change



Baseline



After treatment

- +3: significant improvement
- +2: moderate improvement
- +1: minimal improvement

Clinical response

- 0: no change

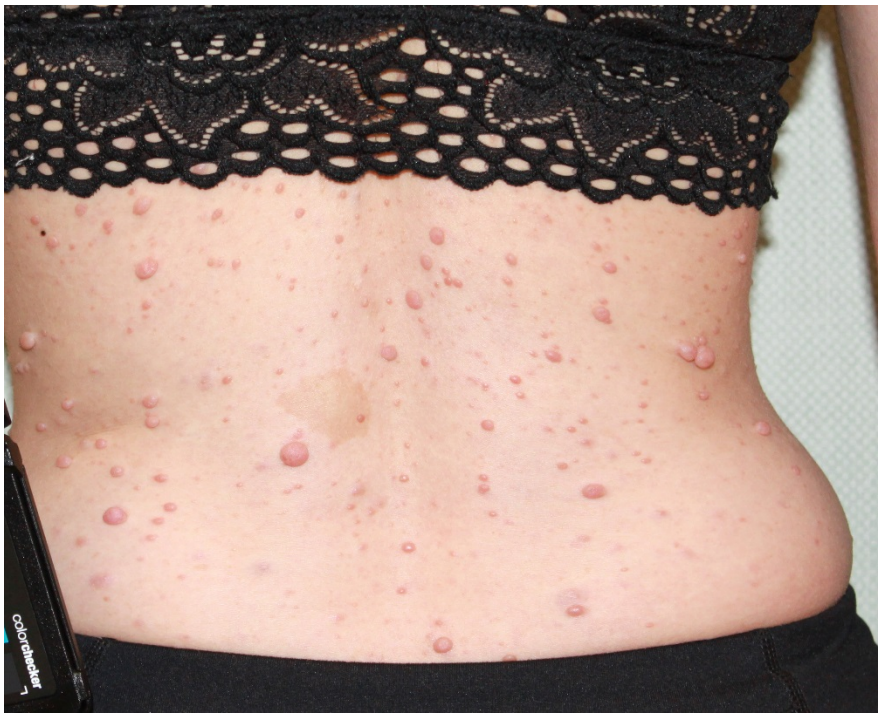
Stable

- -1: minimal worsening
- -2: moderate worsening
- -3: significant worsening

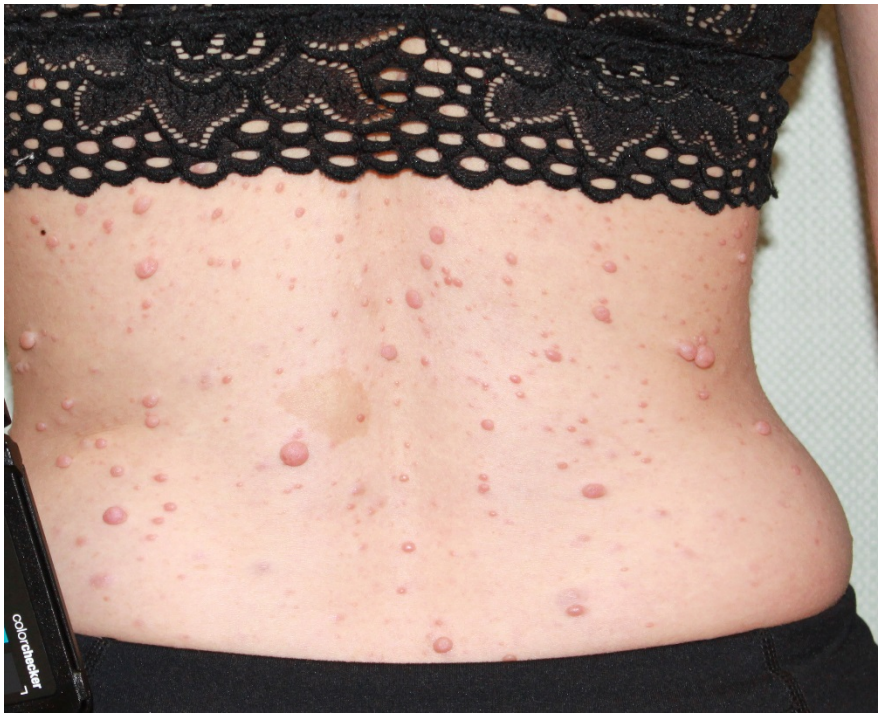
Clinical worsening

shutterstock.com • 1515835526

S3. Simulation of whole body change



Baseline



After treatment

- +3: significant improvement
- +2: moderate improvement
- +1: minimal improvement

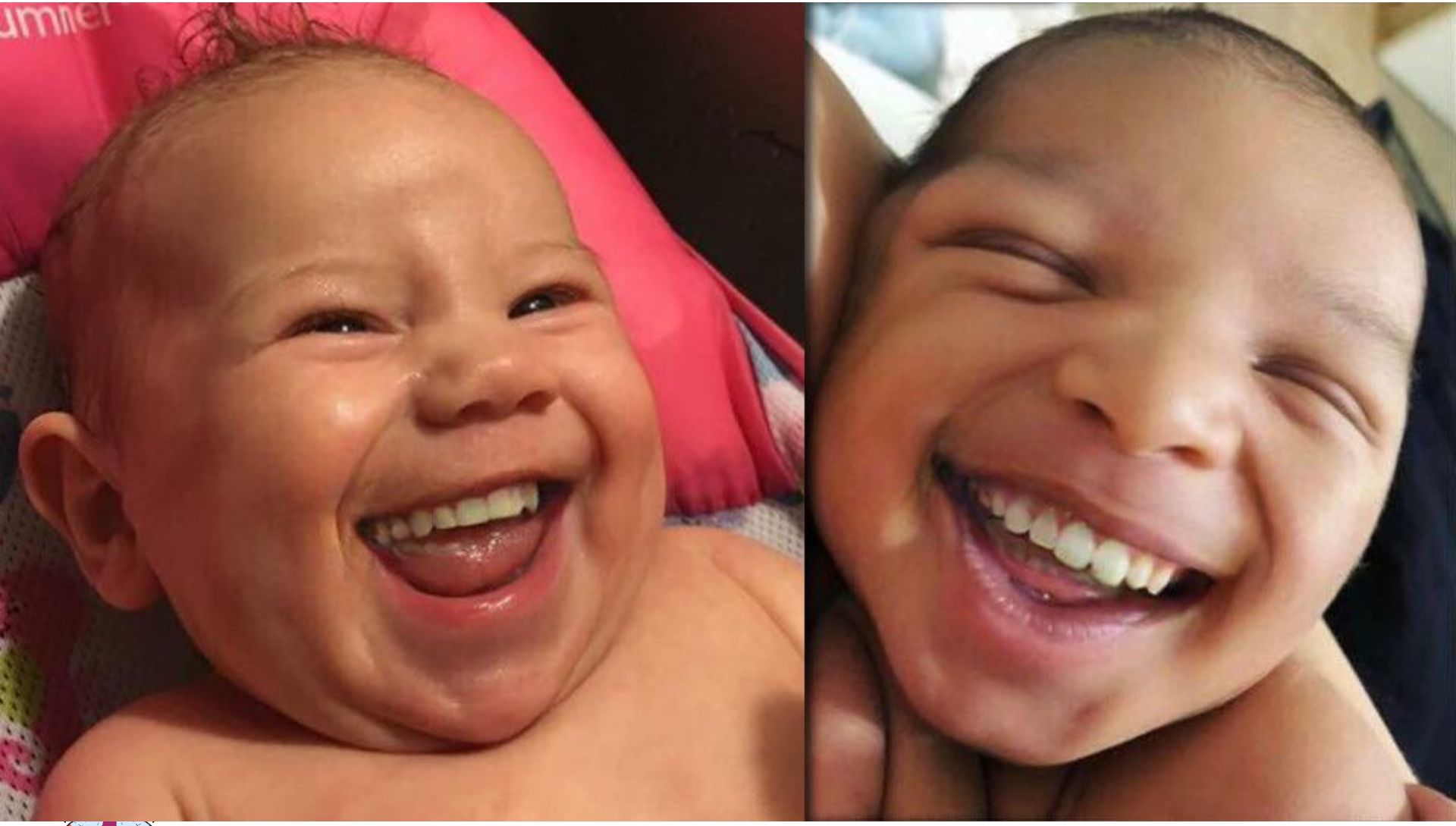
Clinical response

- 0: no change

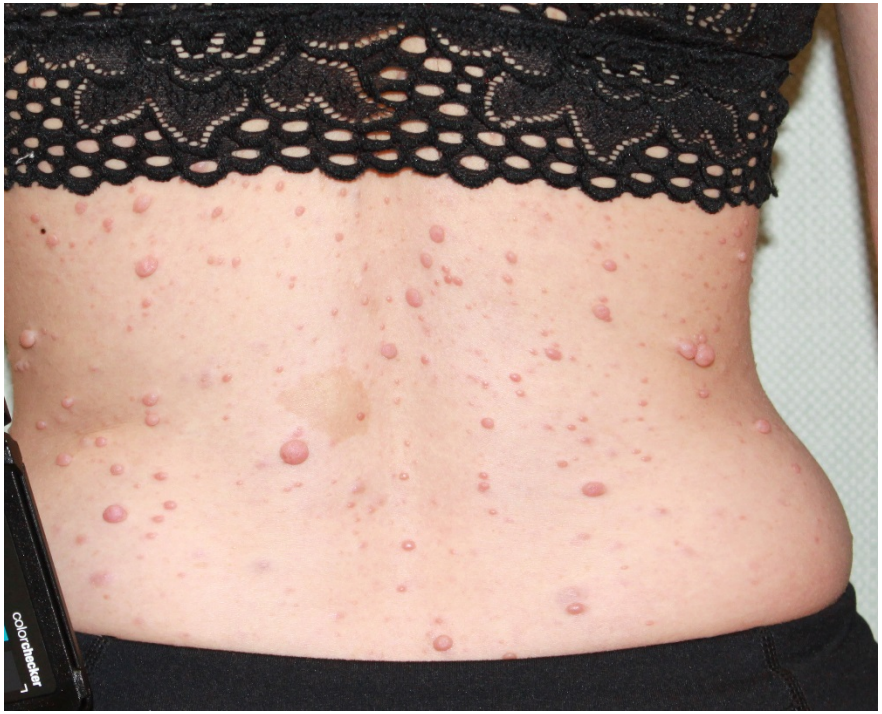
Stable

- -1: minimal worsening
- -2: moderate worsening
- -3: significant worsening

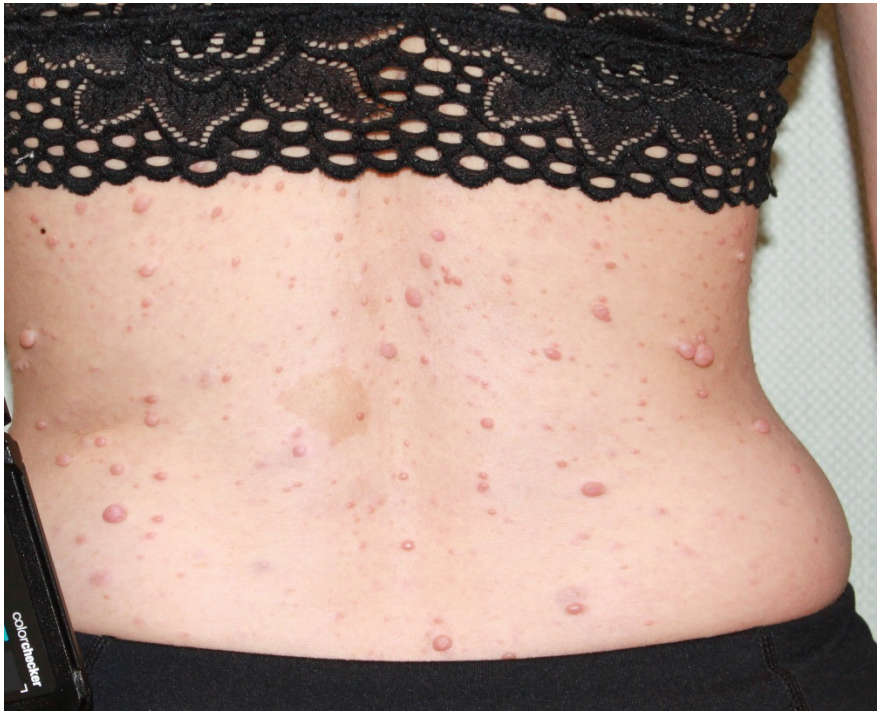
Clinical worsening



S4. Simulation of whole body change



Baseline



After treatment

- +3: significant improvement
- +2: moderate improvement
- +1: minimal improvement

Clinical response

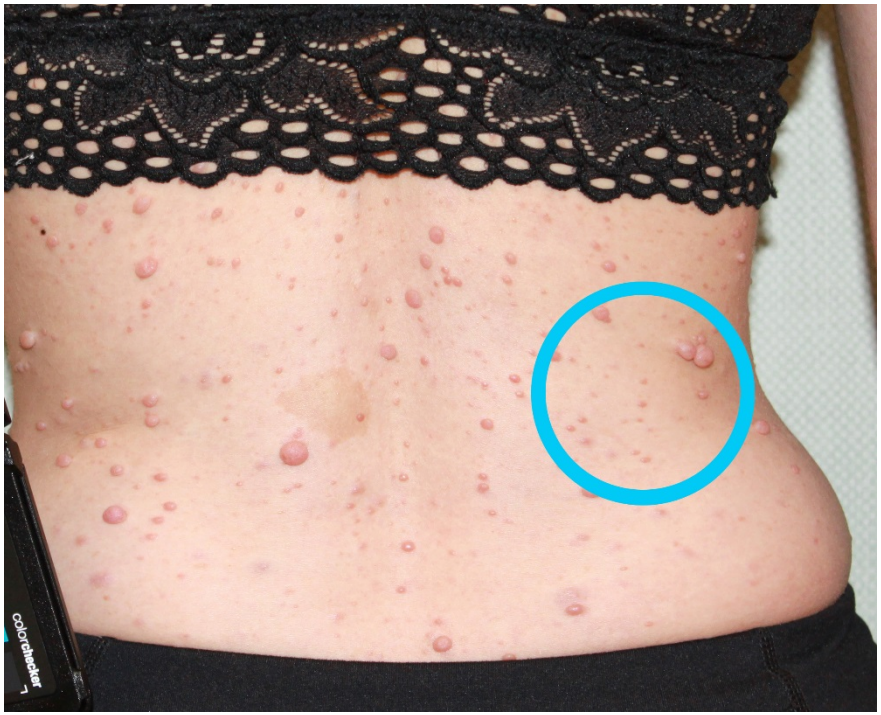
- 0: no change

Stable

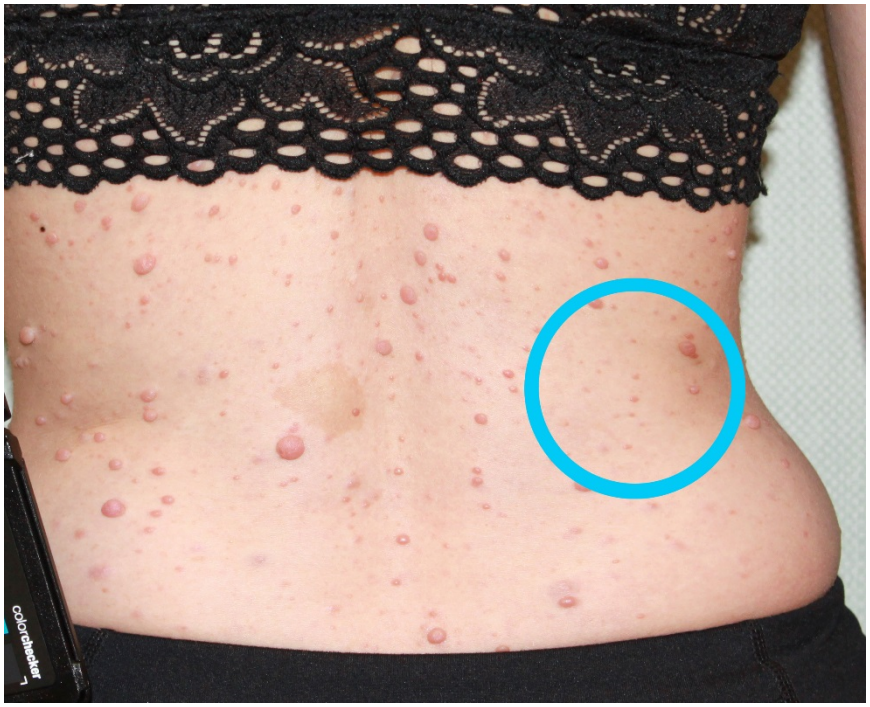
- -1: minimal worsening
- -2: moderate worsening
- -3: significant worsening

Clinical worsening

S5. Simulation of regional change



Baseline



After treatment

- +3: significant improvement
- +2: moderate improvement
- +1: minimal improvement

Clinical response

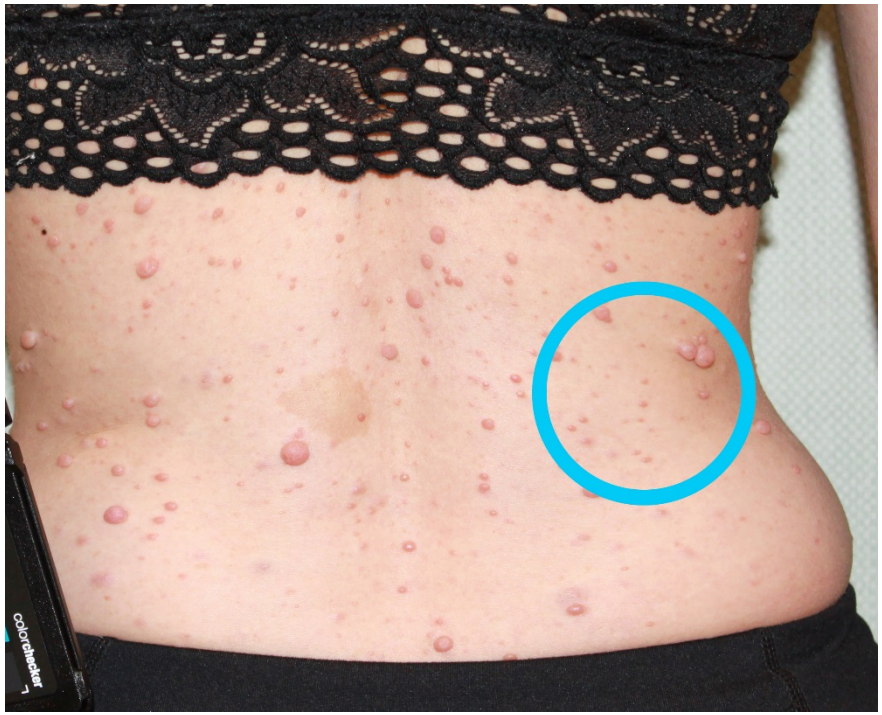
- 0: no change

Stable

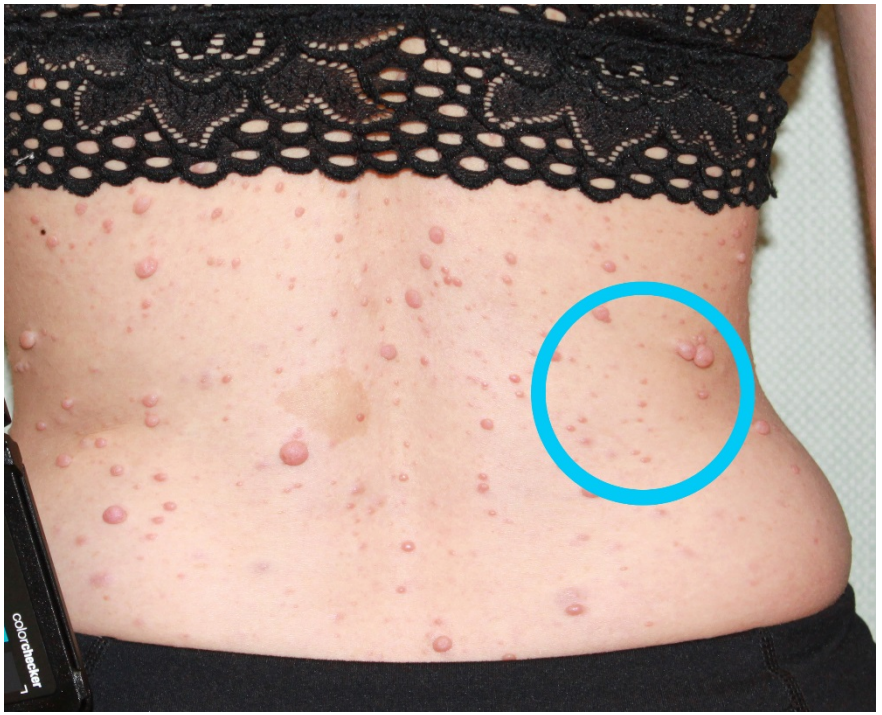
- -1: minimal worsening
- -2: moderate worsening
- -3: significant worsening

Clinical worsening

S6. Simulation of regional change



Baseline



After treatment

- +3: significant improvement
- +2: moderate improvement
- +1: minimal improvement

Clinical response

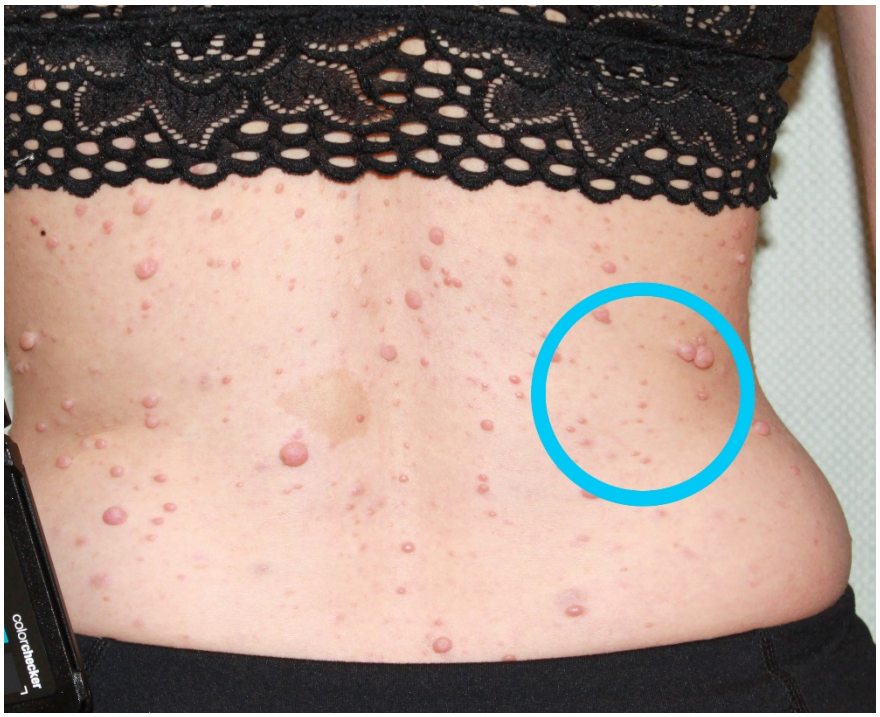
- 0: no change

Stable

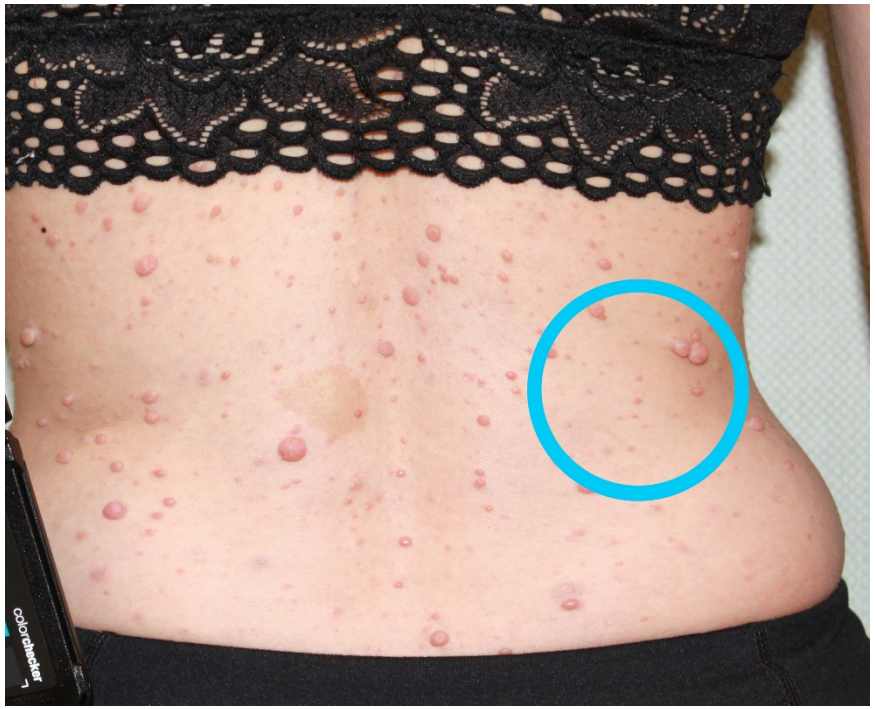
- -1: minimal worsening
- -2: moderate worsening
- -3: significant worsening

Clinical worsening

S7. Simulation of regional change



Baseline



After treatment

- +3: significant improvement
- +2: moderate improvement
- +1: minimal improvement

Clinical response

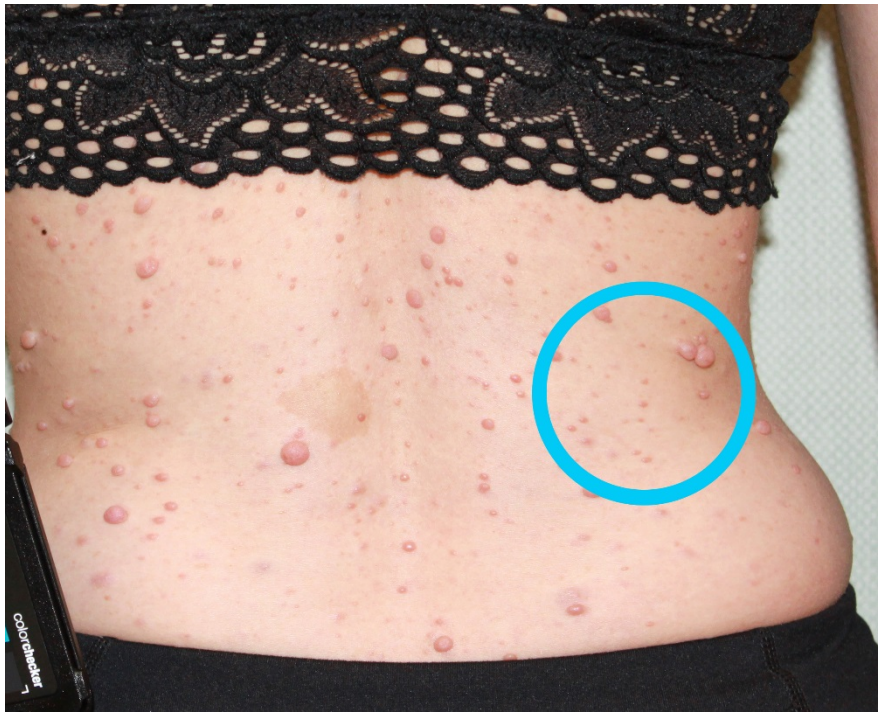
- 0: no change

Stable

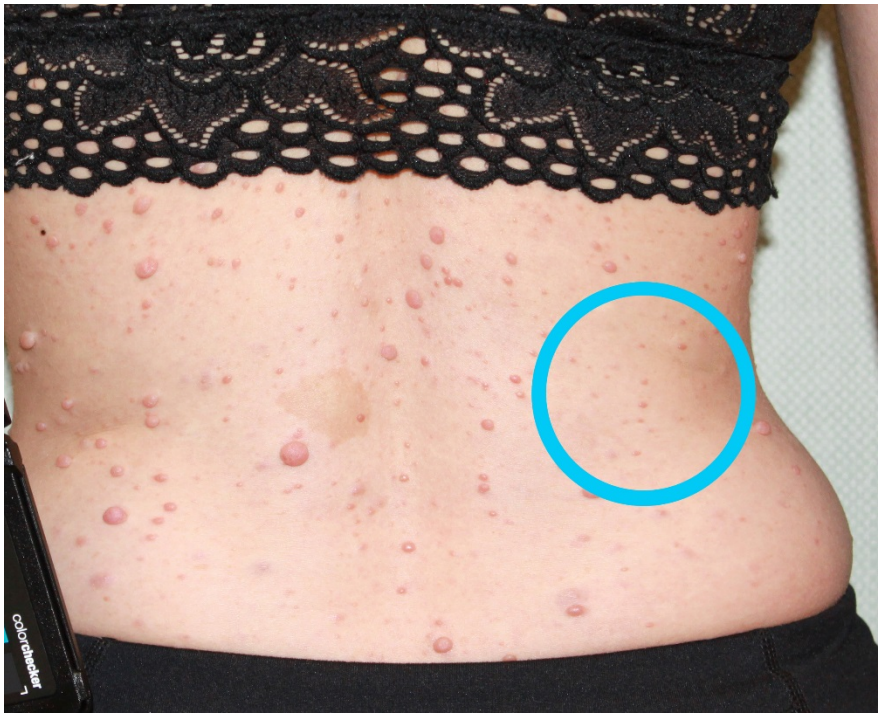
- -1: minimal worsening
- -2: moderate worsening
- -3: significant worsening

Clinical worsening

S8. Simulation of regional change



Baseline



After treatment

- +3: significant improvement
- +2: moderate improvement
- +1: minimal improvement

Clinical response

- 0: no change

Stable

- -1: minimal worsening
- -2: moderate worsening
- -3: significant worsening

Clinical worsening

Next steps

- Complete simulations of 10 subjects
- Assess agreement among large number of patients and physicians using NF registry or other mechanism

Next step: establishing GAC for cNF

- Assemble library of 50 images of cNF with different skin tones and severity
- Digital artist to take each source image and modify
  - Increase tumor size by 25%, 50%, 75%
  - Decrease tumor size by 25%, 50%, 75%
- Recruit online sample to rate photos: baseline with 1 modified image
  - Patients with NF1
  - Caregivers with NF1
  - Investigators who care for patients with NF1
- Rate pairs of images on GAC scale
- Attempt to obtain thousands of responses to establish link between change in tumor size and change in GAC scale

For local treatment: “Compared to baseline, please rate global change in size within the field of treatment”

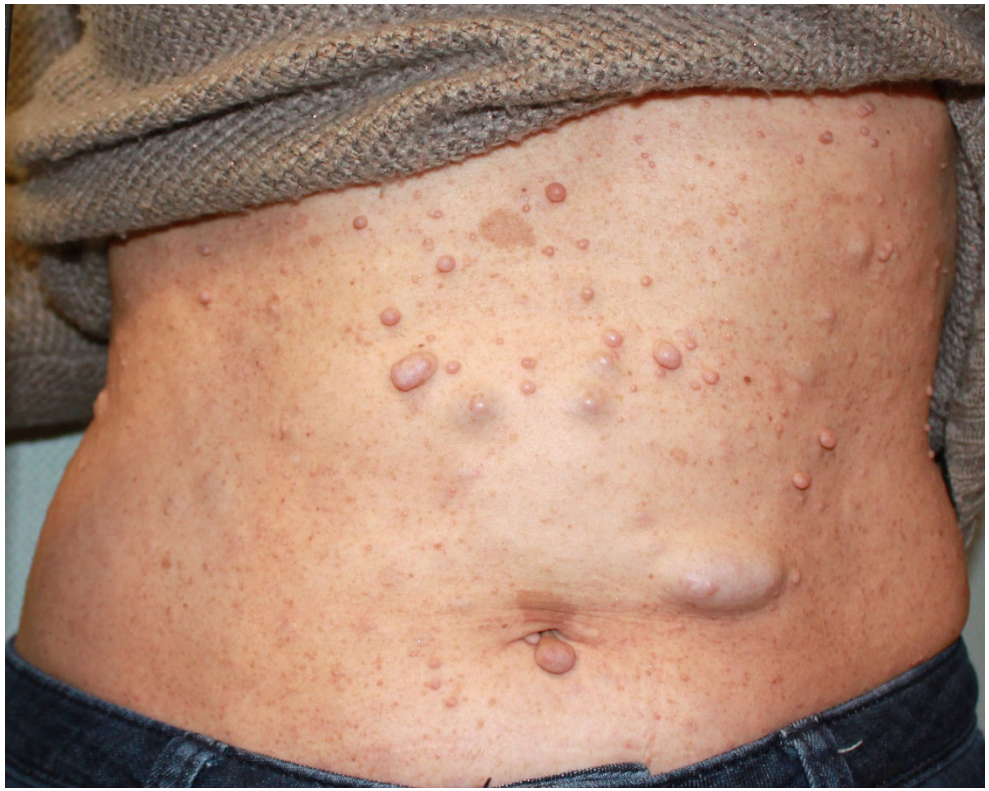
For systemic treatment: “Compared to baseline, please rate global change in size within the field of treatment”

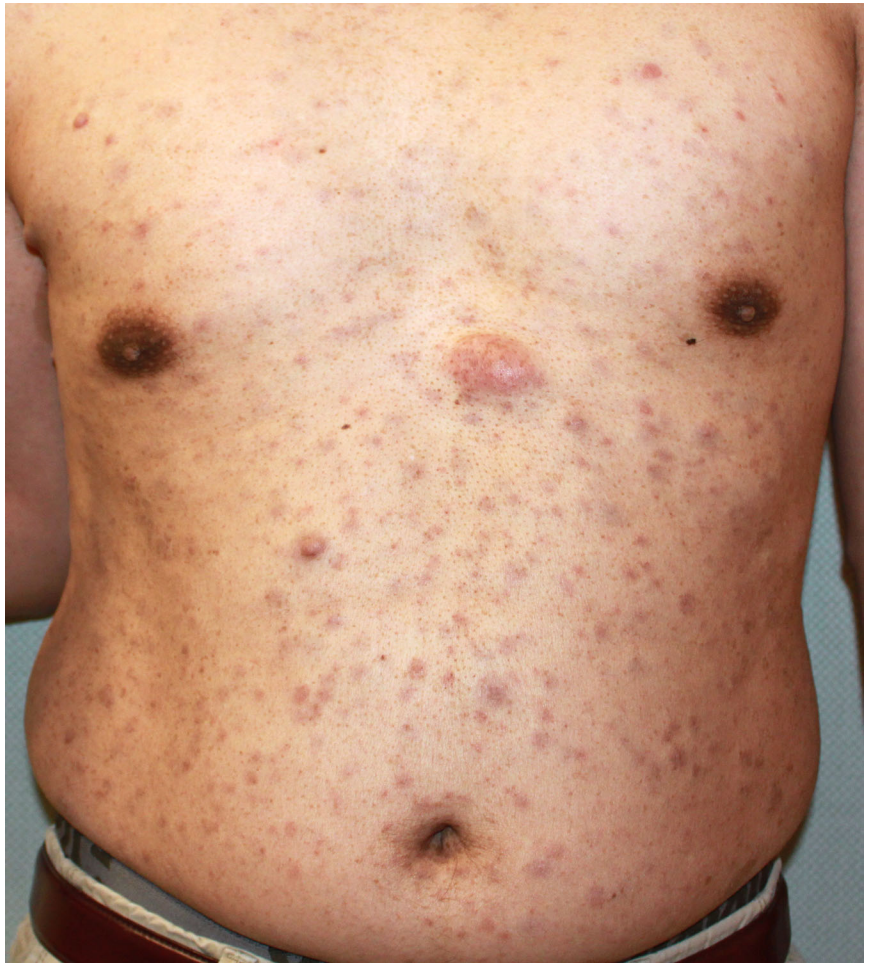
% change compared with baseline (generated)

Global assessment of change

	-75%	-50%	-25%	0%	25%	50%	75%
+3	38	60					
+2	23	20	4				
+1		20	34				
0		2	40				
-1			3				
-2							
-3							

Challenges: phenotypic heterogeneity of cNF





Proposed global assessment of change scale for cNF visibility (color)

- +3: significant improvement
- +2: moderate improvement
- +1: minimal improvement
- 0: no change
- -1: minimal worsening
- -2: moderate worsening
- -3: significant worsening

Clinical response

Stable

Clinical worsening

Questions?