PharmaSUG 2023 - Paper DS-125

Questionnaires in ADaM: ADSF36

Keith Shusterman, Mario Widel, and Megan O'Grady, Reata Pharmaceuticals Inc.

ABSTRACT

The 36-Item Short Form Survey (SF-36) questionnaire is one of the most widely utilized generic health measures in clinical trials. The SF-36 User Manual recommends that population norm-based scores should be used to simplify interpretation of the data. The responses to the SF-36 are utilized to derive a physical component summary, and a mental component summary. Due to norm-based score methods, the process for deriving these two summary values from the raw scores is complicated. It is common for an independent party to derive the summary scores from the raw scores to be included as a source in the SDTM QS domain. However, there may be a compelling reason to derive these values in-house. For example, vendor algorithms are typically proprietary and cannot be incorporated into submission data and documentation. In this case, the process of deriving the summary scores needs to be appropriately documented in ADaM in a clear and traceable way. In this paper, we will outline the process of deriving and documenting in metadata both the physical component and mental component summary scores in a way that is conformant to the ADaM BDS structure, traceable, and analysis-ready.

INTRODUCTION

In 1956, the US congress passed The National Health Survey Act, which provided the legislative authorization to conduct surveys intended to assess the amount, distribution, and effects of illness in the United States. This regulation enabled various organizations to legally develop and conduct health related surveys.

During the 1960's, the NCHS (National Center for Health Statistics), a branch of the US Public Health Service, conducted several National Health Examination Surveys. In the 1970's, these surveys were expanded to include nutrition and its relationship to health status.

In the late 1940's, The RAND corporation was created as a non-profit organization with the goal to "further promote scientific, educational, and charitable purposes, all for the public welfare and security of the United States of America", as expressed in their mission statement.

During the 1980's, RAND conducted "The Medical Outcomes Study: Measures of Quality-of-Life Core Survey (MOS) ", a two-year study of patients with chronic conditions. The questionnaire had 116 items in their core survey covering measures of quality of life such as physical, mental, and general health.

Later, RAND developed the 36-Item Short Form Health Survey (SF-36 v1) as a set of generic, coherent, and easily administered quality-of-life measures.

In the 1990's, the 36-Item Short Form Health Survey (SF-36 v2) was published. This new version improved the instructions and make the layout easier to read. As these questionnaires were being developed, several variations having fewer questions were considered and tested: SF-20, SF-12, SF-6D.

In the late 2010's, while working on a study with the SF-36 questionnaire, Reata decided to derive the mental and physical component summary scores as part of in-house analysis programming without rely. We will describe how we calculated those scores and how we incorporated the derivations in an ADaM dataset we called ADSF36.

ANALYSIS VALUES AND HOW TO GET THEM

The SF-36 questionnaire collects patient responses to 36 questions. We mapped our responses to the QS domain using the CDISC codelist SF363TC (The Short Form 36 Health Survey Standard, US Version 2.0 Questionnaire Test Code) for QSTESTCD and the codelist SF363TN (The Short Form 36 Health Survey Standard, US Version 2.0 Questionnaire Test Name) for QSTEST. The 2022-06-24 version of SDTM controlled terminology contains several SF-36 codelists in addition to supporting various versions of the questionnaire, so use caution when selecting CDISC codelists.

SDTM QS usually involves a numeric result mapped to QSSTRESN, as well as a character decode mapped to QSSTRESC. For example, the first question of "In general, would you say your health is:" can have responses ranging from excellent (QSSTRESN=1, QSSTRESC = EXCELLENT) to poor (QSSTRESN = 5, QSSTRESC = POOR). An example SDTM-annotated CRF for the first question of the questionnaire is shown in Figure 1. We did confirm that the numeric responses in our database were the proper numeric responses for the questionnaire. Caution is advised, as this is not guaranteed to be the case for any given database.

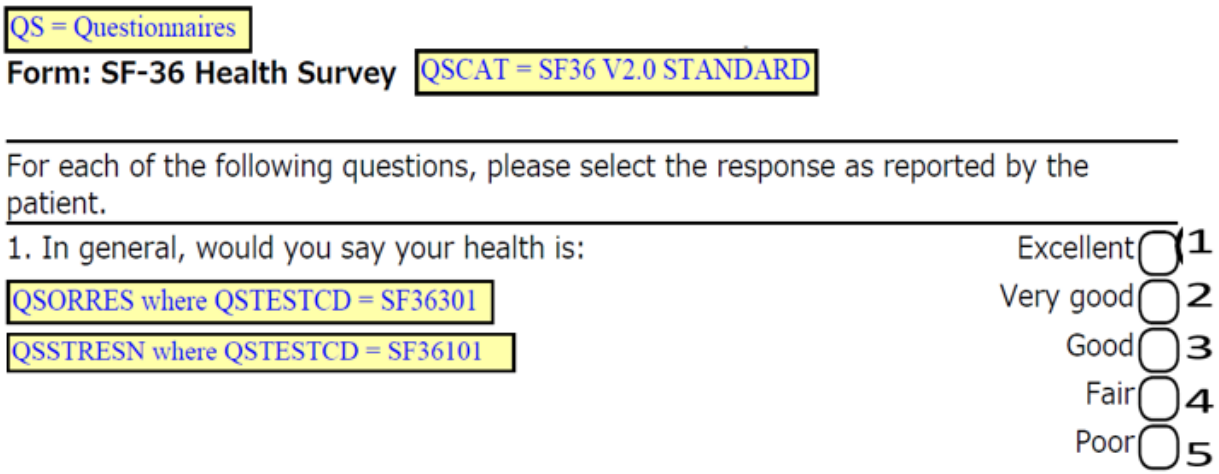

Figure 1: Example of an SDTM-annotated CRF for the first question of the SF-36 questionnaire

The end goal is to derive physical and mental component summary scores. The process we used to derive these two scores from the 36 patient responses involves seven steps that can be each be represented in ADaM for maximum traceability. The seven steps are illustrated in Figure **2**. We will describe how we included parameters supporting each of these steps in an analysis-ready BDS dataset called ADSF36.

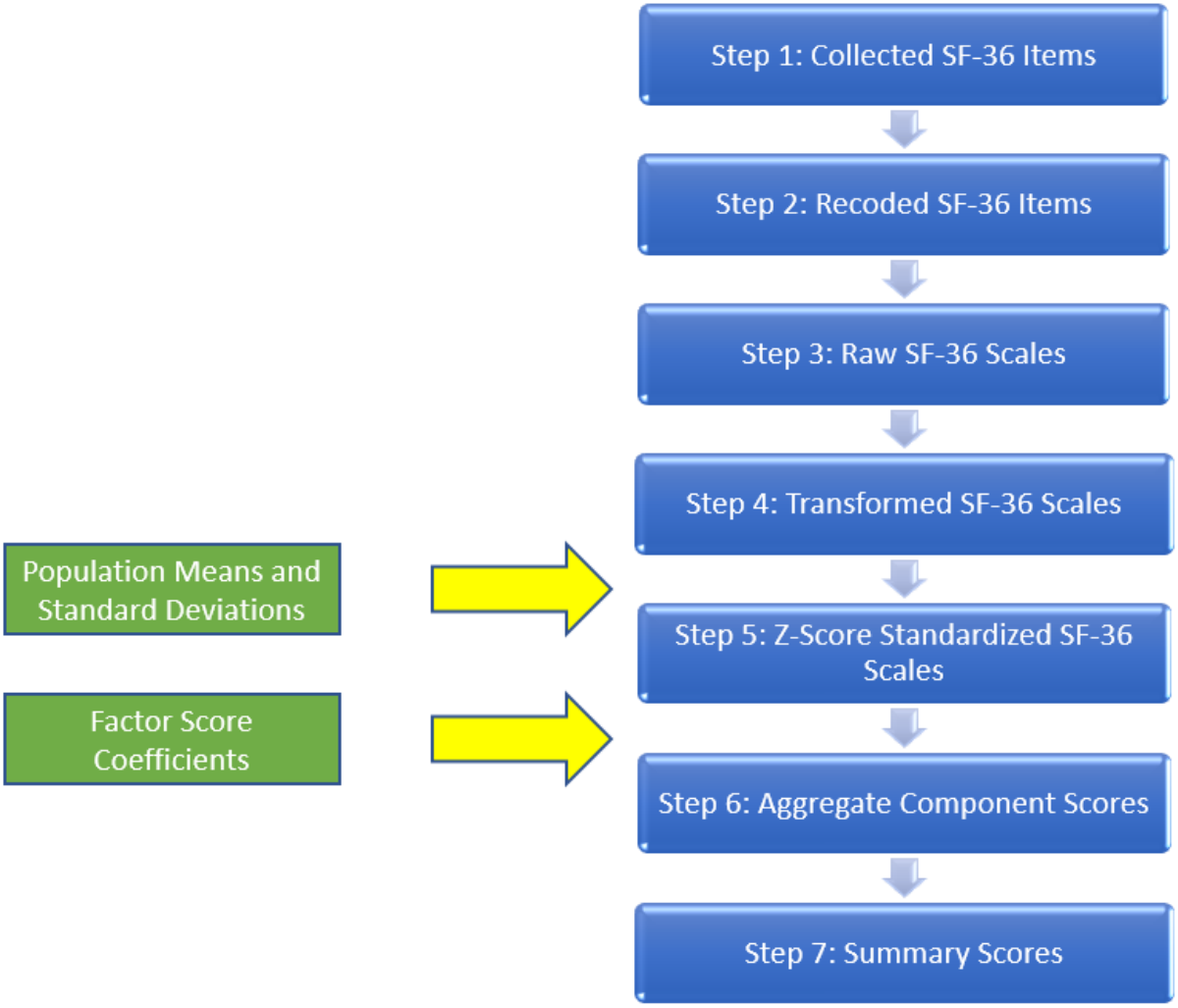

Figure 2: The seven steps of deriving summary scores from collected SF-36 responses.

COLLECTED SF-36 ITEMS (STEP 1)

The first step is to create a set of parameters that represent the collected SF-36 responses. To make the connection as clear as possible between these parameters and the SDTM data, we set PARAMCD and PARAM to match QSTESTCD and QSTEST, respectively. These parameters have a predecessor origin unless record-level imputations or derivations are performed and documented using the DTYPE variable. AVAL is populated from QSSTRESN. There is no analysis need to include the QSSTRESC values in AVALC, but we included them for traceability and to provide additional context. We used PARCAT1 = "Collected SF-36 Responses". Throughout this process, we populated PARCAT1N with the step number to show the sequence in which these sets of parameters are derived. The variables SRCDOM, SRCVAR, and SRCSEQ were utilized to show traceability from QS with SRCDOM = "QS", SRCVAR = "QSSTRESN", and SRCSEQ set to QSSEQ from the associated QS record. An example of data with selected variables and parameters is shown in Table 1, and an example of value level metadata with selected parameters is shown in Table 2. Note that the second question (QSTESTCD=" SF36302") is not used in any analysis and does not need to be included in ADSF36.

STUDYID	USUBJID	QSTESTCD	QSTEST	PARCAT1N	PARCAT1	PARAMCD	PARAM	ADT	AVAL	SRCDOM	SRCVAR	SRCSEQ
ABC	ABC-123	SF36301	SF363-In General You Say Your Health Is	1	Collected SF-36 Responses	SF36301	SF363-In General You Say Your Health Is	01JAN2023	2	QS	QSSTRESN	1
ABC	ABC-123	SF36303A	SF363-Limit Vigorous Activities	1	Collected SF-36 Responses	SF36303A	SF363-Limit Vigorous Activities	01JAN2023	4	QS	QSSTRESN	3
ABC	ABC-123	SF36303B	SF363-Limit Moderate Activities	1	Collected SF-36 Responses	SF36303B	SF363-Limit Moderate Activities	01JAN2023	3	QS	QSSTRESN	4

Table 1: Example data with selected variables and parameters for collected SF-36 responses in ADSF36.

Value Level Metadata -ADSF36 [AVAL]

Variable	Where	Type	Length / Display Format	Controlled Terms or Format	Origin	Derivation/ Comment
AVAL	PARAMCD = "SF36301" (SF363-In General You Say Your Health Is)	float	3		Predecessor	
AVAL	PARAMCD = "SF36303A" (SF363-Limit Vigorous Activities)	float	3		Predecessor	
AVAL	PARAMCD = "SF36303B" (SF363-Limit Moderate Activities)	float	3		Predecessor	

Table 2: Value level metadata with selected parameters for collected SF-36 responses in ADSF36.

RECODED SF-36 ITEMS (STEP 2)

The original numbers recorded in QS require several transformations to support calculation of the summary scores First, we recoded the collected numeric values so that all questions have responses where a higher value indicates a better outcome. For example, the first item of the questionnaire collects a value of 1 for a response of “Excellent”, and a value of 5 for “Poor”. Of the 35 questions used in analysis, we recoded 10 of them as such. We left the responses of the remaining 25 questions unchanged in the recoding process. We still created a new set of 35 “recoded” parameters to simplify selection of recoded records in later steps. We used PARCAT1 = “Recoded SF-36 Responses” for these records. An example of value level metadata with selected parameters is shown in Table 3. Note that Table 3 includes all parameters whose values change in this recoding step, as well as recoding instructions for each parameter. To see more details on these derivations, see Appendix 1.

Value Level Metadata -ADSF36 [AVAL]

Variable	Where	Origin	Derivation/Comment
AVAL	PARAMCD = "SF3601R" (SF3601 Recoded)	Derived	Derived based on AVAL where PARAMCD = "SF36301" for the same date/visit. If 1, set to 5. If 2, set to 4.4. If 3, set to 3.4. If 4, set to 2. If 5, set to 1.
AVAL	PARAMCD = "SF3603AR" (SF3603A Recoded)	Derived	Derived as AVAL where PARAMCD = "SF36303A" for the same date/visit.
AVAL	PARAMCD = "SF3606R" (SF3606 Recoded)	Derived	Set to 6 - AVAL where PARAMCD = "SF36306" for the same date/visit.
AVAL	PARAMCD = "SF3607R" (SF3607 Recoded)	Derived	Derived based on AVAL where PARAMCD = "SF36307" for the same date/visit. If 1, set to 6. If 2, set to 5.4. If 3, set to 4.2. If 4, set to 3.1. If 5, set to 2.2. If 6, set to 1.
AVAL	PARAMCD = "SF3608R" (SF3606 Recoded)	Derived	Derived based on AVAL where PARAMCD = "SF36307" and PARAMCD = "SF36308" for the same date/visit. If there is a response for PARAMCD = "SF36307", then derive as follows: If AVAL = 1 for PARAMCD = "SF36308" and AVAL = 1 for PARAMCD = "SF36307", then set to 6. If AVAL = 1 for PARAMCD = "SF36308" and AVAL is not 1 for PARAMCD = "SF36307", then set to 5. If AVAL = 2 for PARAMCD = "SF36308" then set to 4. If AVAL = 3 for PARAMCD = "SF36308" then set to 3. If AVAL = 4 for PARAMCD = "SF36308" then set to 2. If AVAL = 5 for PARAMCD = "SF36308" then set to 1. If there is no response for PARAMCD = "SF36307", then derived based on AVAL where PARAMCD = "SF36308". If 1 then set to 6. If 2 then set to 4.75. If 3 then set to 3.5. If 4 then set to 2.25. If 5 then set to 1.
AVAL	PARAMCD = "SF3609AR" (SF3609A Recoded)	Derived	Set to 7 - AVAL where PARAMCD = "SF36309A" for the same date/visit.
AVAL	PARAMCD = "SF3609DR" (SF3609D Recoded)	Derived	Set to 7 - AVAL where PARAMCD = "SF36309D" for the same date/visit.
AVAL	PARAMCD = "SF3609ER" (SF3609E Recoded)	Derived	Set to 7 - AVAL where PARAMCD = "SF36309E" for the same date/visit.
AVAL	PARAMCD = "SF3609HR" (SF3609H Recoded)	Derived	Set to 7 - AVAL where PARAMCD = "SF36309H" for the same date/visit.
AVAL	PARAMCD = "SF3611BR" (SF3611B Recoded)	Derived	Set to 6 - AVAL where PARAMCD = "SF36311B" for the same date/visit.
AVAL	PARAMCD = "SF3611DR" (SF3611D Recoded)	Derived	Set to 6 - AVAL where PARAMCD = "SF36311D" for the same date/visit.

Table 3: Value level metadata with selected parameters for recoded SF-36 responses in ADSF36, including all parameters where analysis values change via recoding.

These 35 parameters are used to derive eight scales. These scales are Physical Functioning, Role-Physical, Bodily Pain, General Health, Vitality, Social Functioning, Role-Emotional, and Mental Health. Each recoded parameter contributes to exactly one of these scales. As such, we used PARCAT2 to capture the scale that contains its corresponding parameters to simplify the derivation in the next step. The recoded parameters and their respective PARCAT2 values are shown in Table **4**.

PARCAT2	PARAMCD Values
Physical Functioning	SF3603AR, SF3603BR, SF3603CR, SF3603DR, SF3603ER, SF3603FR, SF3603GR, SF3603HR, SF3603IR, SF3603JR
Role-Physical	SF3604AR, SF3604BR, SF3604CR, SF3604DR
Bodily Pain	SF3607R, SF3608R
General Health	SF3601R, SF3611AR, SF3611BR, SF3611CR, SF3611DR
Vitality	SF3609AR, SF3609ER, SF3609GR, SF3909IR
Social Functioning	SF3606R, SF3610R
Role-Emotional	SF3605AR, SF3605BR, SF3605CR
Mental Health	SF3609BR, SF3609CR, SF3609DR, SF3909FR, SF3609HR

Table **4**: PARCAT2 values for recoded SF36 response parameters

An example of data with selected variables and parameters is shown in Table **5**.

STUDYID	USUBJID	PARCAT1	PARCAT2	PARAMCD	PARAM	AVAL
ABC	ABC-124	Recoded SF-36 Responses	General Health	SF3601R	SF3601 Recoded	2
ABC	ABC-124	Recoded SF-36 Responses	Physical Functioning	SF3603AR	SF3603A Recoded	4
ABC	ABC-124	Recoded SF-36 Responses	Physical Functioning	SF3603BR	SF3603B Recoded	3

Table 5: Example data with selected variables and parameters for recoded SF-36 responses in ADSF36

RAW SF-36 SCALES (STEP 3)

The 35 recoded SF-36 response parameters are now used to derive eight scales. In this step, eight raw scale scores are derived from the 35 recoded SF-36 responses. The raw scales are simply derived as the sum of the recoded responses for each parameter associated with that scale. For these parameters, we used PARCAT1 = “Raw SF-36 Scales”. An example of value level metadata for all eight parameters is shown in Table 6. To see more details on these derivations, see Appendix 1.

Value Level Metadata -ADSF36 [AVAL]

Variable	Where	Origin	Derivation/Comment
AVAL	PARAMCD = "PFRS" (Physical Functioning Raw Scale)	Derived	Set to the sum of AVAL values for parameters at the same date/visit with PARCAT2 = "Physical Functioning"
AVAL	PARAMCD = "RPRS" (Role-Physical Raw Scale)	Derived	Set to the sum of AVAL values for parameters at the same date/visit with PARCAT2 = "Role-Physical"
AVAL	PARAMCD = "BPRS" (Bodily Pain Raw Scale)	Derived	Set to the sum of AVAL values for parameters at the same date/visit with PARCAT2 = "Bodily Pain"
AVAL	PARAMCD = "GHRS" (General Health Raw Scale)	Derived	Set to the sum of AVAL values for parameters at the same date/visit with PARCAT2 = "General Health"
AVAL	PARAMCD = "VTRS" (Vitality Raw Scale)	Derived	Set to the sum of AVAL values for parameters at the same date/visit with PARCAT2 = "Vitality"
AVAL	PARAMCD = "SFRS" (Social Functioning Raw Scale)	Derived	Set to the sum of AVAL values for parameters at the same date/visit with PARCAT2 = "Social Functioning"
AVAL	PARAMCD = "RERS" (Role-Emotional Raw Scale)	Derived	Set to the sum of AVAL values for parameters at the same date/visit with PARCAT2 = "Role-Emotional Functioning"
AVAL	PARAMCD = "MHRS" (Mental Health Raw Scale)	Derived	Set to the sum of AVAL values for parameters at the same date/visit with PARCAT2 = "Mental Health"

Table 6: Value Level Metadata for Raw Scale Scores in ADSF36.

An example of data with selected variables and parameters is shown in Table 7.

STUDYID	USUBJID	PARCAT1N	PARCAT1	PARAMCD	PARAM	ADT	AVAL
ABC	ABC-125	3	Raw SF-36 Scales	PFRS	Physical Functioning Raw Scale	01JAN2023	9
ABC	ABC-125	3	Raw SF-36 Scales	RPRS	Role-Physical Raw Scale	01JAN2023	11
ABC	ABC-125	3	Raw SF-36 Scales	BPRS	Bodily Pain Raw Scale	01JAN2023	2

Table 7: Example data with selected variables and parameters for SF-36 Raw Scale Scores in ADSF36

TRANSFORMED SF-36 SCALES (STEP 4)

Now that we have the raw SF-36 scales, the next step is to transform these raw scales such that the score is on a scale from 0 to 100. We did this by taking the raw score, subtracting from it the minimum possible score for that scale, dividing the difference by the maximum possible score for that scale, and then multiplying by 100. For these parameters, we used PARCAT1 = “Transformed SF-36 Scales”. An example of value level metadata for all eight parameters is shown in Table 8. To see more details on these derivations, see Appendix 1.

Value Level Metadata -ADSF36 [AVAL]

Variable	Where	Origin	Derivation/Comment
AVAL	PARAMCD = "PFTS" (Physical Functioning Transformed Scale)	Derived	Set to (PFRS - 10)/20 * 100, where PFRS = AVAL where PARAMCD = "PFRS" for the same visit/date.
AVAL	PARAMCD = "RPTS" (Role-Physical Transformed Scale)	Derived	Set to (RPRS - 4)/16 * 100, where PFRS = AVAL where PARAMCD = "RPRS" for the same visit/date.
AVAL	PARAMCD = "BPTS" (Bodily Pain Transformed Scale)	Derived	Set to (BPRS - 2)/10 * 100, where PFRS = AVAL where PARAMCD = "BPRS" for the same visit/date.
AVAL	PARAMCD = "GHTS" (General Health Transformed Scale)	Derived	Set to (GHRS - 5)/20 * 100, where PFRS = AVAL where PARAMCD = "GHRS" for the same visit/date.
AVAL	PARAMCD = "VTTS" (Vitality Transformed Scale)	Derived	Set to (VTRS - 6)/16 * 100, where PFRS = AVAL where PARAMCD = "VTRS" for the same visit/date.
AVAL	PARAMCD = "SFTS" (Social Functioning Transformed Scale)	Derived	Set to (SFRS - 2)/8 * 100, where PFRS = AVAL where PARAMCD = "SFRS" for the same visit/date.
AVAL	PARAMCD = "RETS" (Role-Emotional Transformed Scale)	Derived	Set to (RERS - 3)/12 * 100, where PFRS = AVAL where PARAMCD = "RERS" for the same visit/date.
AVAL	PARAMCD = "MHTS" (Mental Health Transformed Scale)	Derived	Set to (MHRS - 7)/20 * 100, where PFRS = AVAL where PARAMCD = "MHRS" for the same visit/date.

Table 8: Value Level Metadata for Transformed Scale Scores in ADSF36.

An example of data with selected variables and parameters is shown in Table 9.

STUDYID	USUBJID	PARCAT1N	PARCAT1	PARAMCD	PARAM	ADT	AVAL
ABC	ABC-126	4	Transformed SF-36 Scales	PFTS	Physical Functioning Transformed Scale	01JAN2023	44
ABC	ABC-126	4	Transformed SF-36 Scales	RPTS	Role-Physical Transformed Scale	01JAN2023	82
ABC	ABC-126	4	Transformed SF-36 Scales	BPTS	Bodily Pain Transformed Scale	01JAN2023	56

Table 9: Example data with selected variables and parameters for SF-36 Transformed Scale Scores in ADSF36

POPULATION STATISTICS AND FACTOR SCORE COEFFICIENTS (NEEDED FOR STEPS 5 AND 6)

The next two steps involve transforming the transformed SF-36 scales using population statistics and factor score coefficients. The statistics and coefficients we used came from John E. Ware's SF-36 user's manual from 1994. These are shown in Table 10.

			Factor Score Coefficients	
SF-36 Scale	Mean	SD	PCS	MCS
PF	84.52404	22.89490	0.42402	-0.22999
RP	81.19907	33.79729	0.35119	-0.12329
BP	75.49196	23.55879	0.31754	-0.09731
GH	72.21316	20.16964	0.24954	-0.01571
VT	61.05453	20.86942	0.02877	0.23534
SF	83.59753	22.37642	-0.00753	0.26876
RE	81.29467	33.02717	-0.19206	0.43407
MH	74.84212	18.01189	-0.22069	0.48581

Table 10: SF-36 Population Mean, Population Standard Deviation, Physical Factor Score Component, and Mental Factor Score Component by Scale (Ware 1994, p. 26).

For each of the eight scales, Table 10 gives a population mean, a population standard deviation, a factor score coefficient for the physical component, and a factor score coefficient for the mental component. We used these items in deriving the next two sets of parameters.

Z-SCORE STANDARDIZED SF-36 SCALES (STEP 5)

Now that we have the eight SF-36 scales transformed into a 0-100 scale, we then used the population means and standard deviations in Table 10 to transform each transformed scale into a z-score. This is done by taking each transformed score, subtracting the population mean from it, and then dividing the difference by the population standard deviation. For these parameters, we used PARCAT1 = "Z-Score Standardized SF-36 Scales". An example of value level metadata for all eight parameters is shown in Table 11. To see more details on these derivations, see Appendix 1.

Value Level Metadata -ADSF36 [AVAL]

Variable	Where	Origin	Derivation/Comment
AVAL	PARAMCD = "PFZS" (Physical Functioning Z-Score)	Derived	Set to (PFTS - 84.52404)/22.89490, where PFRS = AVAL where PARAMCD = "PFTS" for the same visit/date.
AVAL	PARAMCD = "RPZS" (Role-Physical Z-Score)	Derived	Set to (RPTS - 81.19907)/33.79729, where PFRS = AVAL where PARAMCD = "RPTS" for the same visit/date.
AVAL	PARAMCD = "BPZS" (Bodily Pain Z-Score)	Derived	Set to (BPTS - 75.49196)/23.55879, where PFRS = AVAL where PARAMCD = "BPTS" for the same visit/date.
AVAL	PARAMCD = "GHZS" (General Health Z-Score)	Derived	Set to (GHTS - 72.21316)/20.16964, where PFRS = AVAL where PARAMCD = "GHTS" for the same visit/date.
AVAL	PARAMCD = "VTZS" (Vitality Z-Score)	Derived	Set to (VTTS - 61.05453)/20.86942, where PFRS = AVAL where PARAMCD = "VTTS" for the same visit/date.
AVAL	PARAMCD = "SFZS" (Social Functioning Z-Score)	Derived	Set to (SFTS - 83.59753)/22.37642, where PFRS = AVAL where PARAMCD = "SFTS" for the same visit/date.
AVAL	PARAMCD = "REZS" (Role-Emotional Z-Score)	Derived	Set to (RETS - 81.29467)/33.02717, where PFRS = AVAL where PARAMCD = "RETS" for the same visit/date.
AVAL	PARAMCD = "MHZS" (Mental Health Z-Score)	Derived	Set to (MHTS - 74.84242)/18.01189, where PFRS = AVAL where PARAMCD = "MHTS" for the same visit/date.

Table 11: Value Level Metadata for Z-Score Standardized Scale Scores in ADSF36.

An example of data with selected variables and parameters is shown in Table 12.

STUDYID	USUBJID	PARCAT1N	PARCAT1	PARAMCD	PARAM	ADT	AVAL
ABC	ABC-127	5	Z-Score Standardized SF-36 Scales	PFZS	Physical Functioning Z-Score	01JAN2023	0.345
ABC	ABC-127	5	Z-Score Standardized SF-36 Scales	RPZS	Role-Physical Transformed Z-Score	01JAN2023	1.1154
ABC	ABC-127	5	Z-Score Standardized SF-36 Scales	BPZS	Bodily Pain Transformed Z-Score	01JAN2023	-0.886

Table 12: Example data with selected variables and parameters for Z-Score Standardized SF-36 Scale Scores in ADSF36

AGGREGATE COMPONENT SCORES (STEP 6)

Once we have the standardized z-scores for each scale, we can then derive aggregate component scores using the factor score coefficients in Table 10. In this step, we went from eight parameters to two parameters. All eight of the z-score standardized scales are used in conjunction with the factor score coefficients to derive a physical aggregate component score and a mental aggregate component score.

To derive the physical aggregate component score, each z-score from the previous step is multiplied by the associated coefficient in the "PCS" column, and then all eight results are summed together to create a single physical aggregate component score.

To derive the mental aggregate component score, each z-score from the previous step is multiplied by the associated coefficient in the "MCS" column, and then all eight results are summed together to create a single mental aggregate component score.

For these parameters, we used PARCAT1 = "Aggregate Component Scores". An example of value level metadata for both parameters is shown in Table 13. To see more details on these derivations, see Appendix 1.

Value Level Metadata -ADSF36 [AVAL]

Variable	Where	Origin	Derivation/Comment
AVAL	PARAMCD = "PACS" (Physical Aggregate Component Score)	Derived	Set to (PFZS x 0.42402) + (RPZS x 0.35119) + (BPZS x 0.31754) + (GHZS x 0.24954) + (VTZS x 0.02877) + (SFZS x -0.00753) + (REZS x -0.19206) + (MHZS x -0.22069), where PFZS = AVAL where PARAMCD = "PFSZ" for the same visit/date, and the other terms are defined similarly.
AVAL	PARAMCD = "MACS" (Mental Aggregate Component Score)	Derived	Set to (PFZS x -0.22999) + (RPZS x -0.12329) + (BPZS x -0.09731) + (GHZS x -0.01571) + (VTZS x 0.23534) + (SFZS x 0.26876) + (REZS x 0.43407) + (MHZS x 0.48581), where PFZS = AVAL where PARAMCD = "PFSZ" for the same visit/date, and the other terms are defined similarly.

Table 13: Value Level Metadata for Z-Score Standardized Scale Scores in ADSF36.

An example of data with selected variables and parameters is shown in Table 14.

STUDYID	USUBJID	PARCAT1N	PARCAT1	PARAMCD	PARAM	ADT	AVAL
ABC	ABC-128	6	Aggregate Component Scores	PACG	Physical Aggregate Component Score	01JAN2023	12.331
ABC	ABC-128	6	Aggregate Component Scores	MACG	Mental Aggregate Component Score	01JAN2023	21.887

Table 14: Example data with selected variables and parameters for Aggregate Component Scores in ADSF36

SUMMARY SCORES (STEP 7)

Lastly, we transform the aggregate component scores into the final component summary scores by applying a T-score transformation. This is done by multiplying the aggregate score by 10, and then adding 50 to the product. For these parameters, we used PARCAT1 = “Summary Scores”. An example of value level metadata for both parameters is shown in Table 15. To see more details on these derivations, see Appendix 1.

Value Level Metadata -ADSF36 [AVAL]

Variable	Where	Origin	Derivation/Comment
AVAL	PARAMCD = "PCS" (Physical Component Score)	Derived	Set to 50 + (PACS * 10), where PACS = AVAL where PARAMCD = "PACS" for the same visit/date.
AVAL	PARAMCD = "MCS" (Mental Component Score)	Derived	Set to 50 + (MACS * 10), where PACS = AVAL where PARAMCD = "MACS" for the same visit/date.

Table 15: Value Level Metadata for Summary Scores in ADSF36.

An example of data with selected variables and parameters is shown in Table **16**.

STUDYID	USUBJID	PARCAT1N	PARCAT1	PARAMCD	PARAM	ADT	AVAL
ABC	ABC-129	7	Summary Scores	PCG	Physical Component Score	01JAN2023	65.448
ABC	ABC-129	7	Summary Scores	MCG	Mental Component Score	01JAN2023	72.667

Table 16**: Example data with selected variables and parameters for Summary Scores in ADSF36**

CONCLUSION

Deriving the physical and mental component scores from the original 36 questionnaire responses is admittedly a long and tedious task. Relying on a vendor to derive these scores can avoid this tedium; but deriving them in house can save time and money in the long run, especially if these scores need to be derived for multiple cuts of data. And perhaps more importantly, deriving scores in-house allows for full visibility into the derivations and unlocks additional, more granular, ADaM modeling options. If it is desired to derive these scores in-house without relying on a vendor, then the derivation can be performed

in a way such that the resulting ADaM dataset is both traceable and analysis-ready. With this approach, a single response-set of 36 records does result in a total of 98 ADaM records, only two of which are analyzed. However, this approach does make every step of the derivation transparent and easy to follow for any reviewer, including internal medical reviewers as well as reviewers from government agencies. It is likely possible to combine steps of this derivation to reduce the number of parameters, but that comes at the cost of traceability.

REFERENCES

Ware JE, Kosinski M, & Keller SD. *SF-36 Physical and Mental Health Summary Scales: A User's Manual.* Boston, MA: Health Assessment Lab, 1994. https://www.researchgate.net/profile/John-Ware-6/publication/292390260_SF36_Physical_and_Mental_Health_Summary_Scales_a_User%27s_Manual/links/5af580264585157136caee31/SF-36-Physical-and-Mental-Health-Summary-Scales-a-Users-Manual.pdf

Medical Outcomes Trust. *How To Score the SF-36 Short-Form Health Survey.* Boston, MA: Medical Outcomes Trust, 1991.

Medical Outcomes Trust. *Scoring Exercise for the SF-36 Health Survey, 2nd edition.* Boston, MA: Medical Outcomes Trust, 1994.

Ware JE, Snow KK, Kosinski M, & Gandek B. *SF-36 Health Survey Manual and Interpretation Guide*. Boston, MA: The Health Institute, New England Medical Center, 1993.

Anne-Sophie Bekx, Éanna Kiely. *Handling non-standardized questionnaires*. Phuse 2014. https://www.lexjansen.com/phuse/2014/cd/CD05.pdf

Karin LaPann, Terek Peterson. *Challenges of Processing Questionnaire Data from Collection to SDTM to ADaM and Solutions using SAS®*. PharmaSUG 2014. https://www.lexjansen.com/pharmasug/2014/DS/PharmaSUG-2014-DS08.pdf

Nancy Brucken, Karin LaPann. ADQRS: *Basic Principles for Building Questionnaire, Rating and Scale Analysis Datasets.* PharmaSUG 2017. https://www.lexjansen.com/pharmasug/2017/DS/PharmaSUG-2017-DS24.pdf

Kathryn Valdés. *Missing by Design: Questionnaire Skip Logic*. WUSS 2008. https://www.lexjansen.com/wuss/2008/tut/tut11.pdf

Medical Outcomes Study Questionnaire Short Form 36 Health Survey (SF-36). https://www.brandeis.edu/roybal/docs/SF-36_website_PDF.pdf.

National Health and Nutrition Examination Survey https://www.cdc.gov/nchs/nhanes/history.html

Public Law 652, 1956. *National Health Survey Act*. https://www.govinfo.gov/content/pkg/STATUTE-70/pdf/STATUTE-70-Pg489-2.pdf

Analysis Data Model (ADaM) and Implementation Guide, v.2.1, IG v.1.1. www.cdisc.org

Study Data Tabulation Model (SDTM) and Implementation Guide, v.1.4, IG v.3.2. www.cdisc.org

CDISC foundational standards QRS. https://www.cdisc.org/standards/foundational/qrs. www.cdisc.org

FURTHER READING

Anita Stewart, John E. Ware, editors. Measuring Functioning and Well-Being. The Medical Outcomes Study Approach https://www.rand.org/pubs/commercial_books/CB361.html

J E Ware Jr., C D Sherbourne. *The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection*. https://pubmed.ncbi.nlm.nih.gov/1593914/

Maglinte GA, Hays RD & Kaplan RM. *US general population norms for telephone administration of the SF-36v2.* https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3582698/

ACKNOWLEDGMENTS

We would like to thank Steve Kirby, Rex Tungala, and Angie Goldsberry for their support and guidance in authoring this paper.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Keith Shusterman
Reata Pharmaceuticals, Inc.
keith.shusterman@reatapharma.com

APPENDIX 1

Value-level derivations for selected ADSF36 parameters:

Variable	Where	Origin	Derivation/Comment
AVAL	PARAMCD = "SF36301" (SF363-In General You Say Your Health Is)	Predecessor	
AVAL	PARAMCD = "SF36303A" (SF363-Limit Vigorous Activities)	Predecessor	
AVAL	PARAMCD = "SF36303B" (SF363-Limit Moderate Activities)	Predecessor	
AVAL	PARAMCD = "SF3601R" (SF3601 Recoded)	Derived	Derived based on AVAL where PARAMCD = "SF36301" for the same date/visit. If 1, set to 5. If 2, set to 4.4. If 3, set to 3.4. If 4, set to 2. If 5, set to 1.
AVAL	PARAMCD = "SF3603AR" (SF3603A Recoded)	Derived	Derived as AVAL where PARAMCD = "SF36303A" for the same date/visit.
AVAL	PARAMCD = "SF3606R" (SF3606 Recoded)	Derived	Set to 6 - AVAL where PARAMCD = "SF36306" for the same date/visit.
AVAL	PARAMCD = "SF3607R" (SF3607 Recoded)	Derived	Derived based on AVAL where PARAMCD = "SF36307" for the same date/visit. If 1, set to 6. If 2, set to 5.4. If 3, set to 4.2. If 4, set to 3.1. If 5, set to 2.2. If 6, set to 1.

AVAL	PARAMCD = "SF3608R" (SF3606 Recoded)	Derived	Derived based on AVAL where PARAMCD = "SF36307" and PARAMCD = "SF36308" for the same date/visit. If there is a response for PARAMCD = "SF36307", then derive as follows: If AVAL = 1 for PARAMCD = "SF36308" and AVAL = 1 for PARAMCD = "SF36307", then set to 6. If AVAL = 1 for PARAMCD = "SF36308" and AVAL is not 1 for PARAMCD = "SF36307", then set to 5. If AVAL = 2 for PARAMCD = "SF36308" then set to 4. If AVAL = 3 for PARAMCD = "SF36308" then set to 3. If AVAL = 4 for PARAMCD = "SF36308" then set to 2. If AVAL = 5 for PARAMCD = 'SF36308" then set to 1. If there is no response for PARAMCD = "SF36307", then derived based on AVAL where PARAMCD = "SF36308". If 1 then set to 6. If 2 then set to 4.75. If 3 then set to 3.5. If 4 then set to 2.25. If 5 then set to 1.
AVAL	PARAMCD = "SF3609AR" (SF3609A Recoded)	Derived	Set to 7 - AVAL where PARAMCD = "SF36309A" for the same date/visit.
AVAL	PARAMCD = "SF3609DR" (SF3609D Recoded)	Derived	Set to 7 - AVAL where PARAMCD = "SF36309D" for the same date/visit.
AVAL	PARAMCD = "SF3609ER" (SF3609E Recoded)	Derived	Set to 7 - AVAL where PARAMCD = "SF36309E" for the same date/visit.
AVAL	PARAMCD = "SF3609HR" (SF3609H Recoded)	Derived	Set to 7 - AVAL where PARAMCD = "SF36309H" for the same date/visit.
AVAL	PARAMCD = "SF3611BR" (SF3611B Recoded)	Derived	Set to 6 - AVAL where PARAMCD = "SF36311B" for the same date/visit.
AVAL	PARAMCD = "SF3611DR" (SF3611D Recoded)	Derived	Set to 6 - AVAL where PARAMCD = "SF36311D" for the same date/visit.
AVAL	PARAMCD = "PFRS" (Physical Functioning Raw Scale)	Derived	Set to the sum of AVAL values for parameters at the same date/visit with PARCAT2 = "Physical Functioning"
AVAL	PARAMCD = "RPRS" (Role-Physical Raw Scale)	Derived	Set to the sum of AVAL values for parameters at the same date/visit with PARCAT2 = "Role-Physical"
AVAL	PARAMCD = "BPRS" (Bodily Pain Raw Scale)	Derived	Set to the sum of AVAL values for parameters at the same date/visit with PARCAT2 = "Bodily Pain"
AVAL	PARAMCD = "GHRS" (General Health Raw Scale)	Derived	Set to the sum of AVAL values for parameters at the same date/visit with PARCAT2 = "General Health"
AVAL	PARAMCD = "VTRS" (Vitality Raw Scale)	Derived	Set to the sum of AVAL values for parameters at the same date/visit with PARCAT2 = "Vitality"
AVAL	PARAMCD = "SFRS" (Social Functioning Raw Scale)	Derived	Set to the sum of AVAL values for parameters at the same date/visit with PARCAT2 = "Social Functioning"
AVAL	PARAMCD = "RERS" (Role-Emotional Raw Scale)	Derived	Set to the sum of AVAL values for parameters at the same date/visit with PARCAT2 = "Role-Emotional Functioning"
AVAL	PARAMCD = "MHRS" (Mental Health Raw Scale)	Derived	Set to the sum of AVAL values for parameters at the same date/visit with PARCAT2 = "Mental Health"
AVAL	PARAMCD = "PFTS" (Physical Functioning Transformed Scale)	Derived	Set to (PFRS - 10)/20 * 100, where PFRS = AVAL where PARAMCD = "PFRS" for the same visit/date.
AVAL	PARAMCD = "RPTS" (Role-Physical Transformed Scale)	Derived	Set to (RPRS - 4)/16 * 100, where PFRS = AVAL where PARAMCD = "RPRS" for the same visit/date.
AVAL	PARAMCD = "BPTS" (Bodily Pain Transformed Scale)	Derived	Set to (BPRS - 2)/10 * 100, where PFRS = AVAL where PARAMCD = "BPRS" for the same visit/date.
AVAL	PARAMCD = "GHTS" (General Health Transformed Scale)	Derived	Set to (GHRS - 5)/20 * 100, where PFRS = AVAL where PARAMCD = "GHRS" for the same visit/date.

AVAL	PARAMCD = "VTTS" (Vitality Transformed Scale)	Derived	Set to (VTRS - 6)/16 * 100, where PFRS = AVAL where PARAMCD = "VTRS" for the same visit/date.
AVAL	PARAMCD = "SFTS" (Social Functioning Transformed Scale)	Derived	Set to (SFRS - 2)/8 * 100, where PFRS = AVAL where PARAMCD = "SFRS" for the same visit/date.
AVAL	PARAMCD = "RETS" (Role-Emotional Transformed Scale)	Derived	Set to (RERS - 3)/12 * 100, where PFRS = AVAL where PARAMCD = "RERS" for the same visit/date.
AVAL	PARAMCD = "MHTS" (Mental Health Transformed Scale)	Derived	Set to (MHRS - 7)/20 * 100, where PFRS = AVAL where PARAMCD = "MHRS" for the same visit/date.
AVAL	PARAMCD = "PFZS" (Physical Functioning Z-Score)	Derived	Set to (PFTS - 84.52404)/22.89490, where PFRS = AVAL where PARAMCD = "PFTS" for the same visit/date.
AVAL	PARAMCD = "RPZS" (Role-Physical Z-Score)	Derived	Set to (RPTS - 81.19907)/33.79729, where PFRS = AVAL where PARAMCD = "RPTS" for the same visit/date.
AVAL	PARAMCD = "BPZS" (Bodily Pain Z-Score)	Derived	Set to (BPTS - 75.49196)/23.55879, where PFRS = AVAL where PARAMCD = "BPTS" for the same visit/date.
AVAL	PARAMCD = "GHZS" (General Health Z-Score)	Derived	Set to (GHTS - 72.21316)/20.16964, where PFRS = AVAL where PARAMCD = "GHTS" for the same visit/date.
AVAL	PARAMCD = "VTZS" (Vitality Z-Score)	Derived	Set to (VTTS - 61.05453)/20.86942, where PFRS = AVAL where PARAMCD = "VTTS" for the same visit/date.
AVAL	PARAMCD = "SFZS" (Social Functioning Z-Score)	Derived	Set to (SFTS - 83.59753)/22.37642, where PFRS = AVAL where PARAMCD = "SFTS" for the same visit/date.
AVAL	PARAMCD = "REZS" (Role-Emotional Z-Score)	Derived	Set to (RETS - 81.29467)/33.02717, where PFRS = AVAL where PARAMCD = "RETS" for the same visit/date.
AVAL	PARAMCD = "MHZS" (Mental Health Z-Score)	Derived	Set to (MHTS - 74.84242)/18.01189, where PFRS = AVAL where PARAMCD = "MHTS" for the same visit/date.
AVAL	PARAMCD = "PACS" (Physical Aggregate Component Score)	Derived	Set to (PFZS x 0.42402) + (RPZS x 0.35119) + (BPZS x 0.31754) + (GHZS x 0.24954) + (VTZS x 0.02877) + (SFZS x -0.00753) + (REZS x -0.19206) + (MHZS x -0.22069), where PFZS = AVAL where PARAMCD = "PFSZ" for the same visit/date, and the other terms are defined similarly.
AVAL	PARAMCD = "MACS" (Mental Aggregate Component Score)	Derived	Set to (PFZS x -0.22999) + (RPZS x -0.12329) + (BPZS x -0.09731) + (GHZS x -0.01571) + (VTZS x 0.23534) + (SFZS x 0.26876) + (REZS x 0.43407) + (MHZS x 0.48581), where PFZS = AVAL where PARAMCD = "PFSZ" for the same visit/date, and the other terms are defined similarly.
AVAL	PARAMCD = "PCS" (Physical Component Score)	Derived	Set to 50 + (PACS * 10), where PACS = AVAL where PARAMCD = "PACS" for the same visit/date.
AVAL	PARAMCD = "MCS" (Mental Component Score)	Derived	Set to 50 + (MACS * 10), where PACS = AVAL where PARAMCD = "MACS" for the same visit/date.