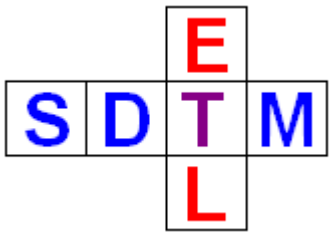
SDTM-ETL™



New features in version 1.5

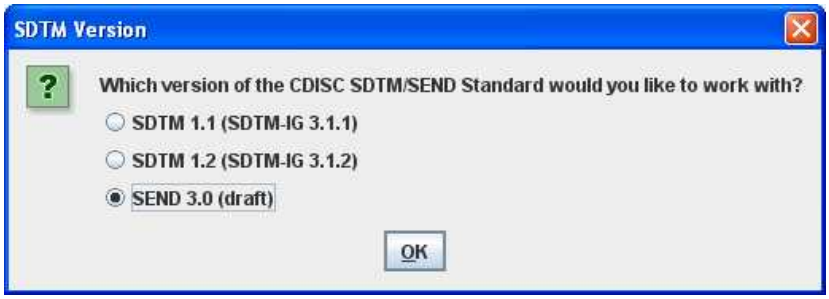
Author: Jozef Aerts – XML4Pharma
December 2010

Table of Contents

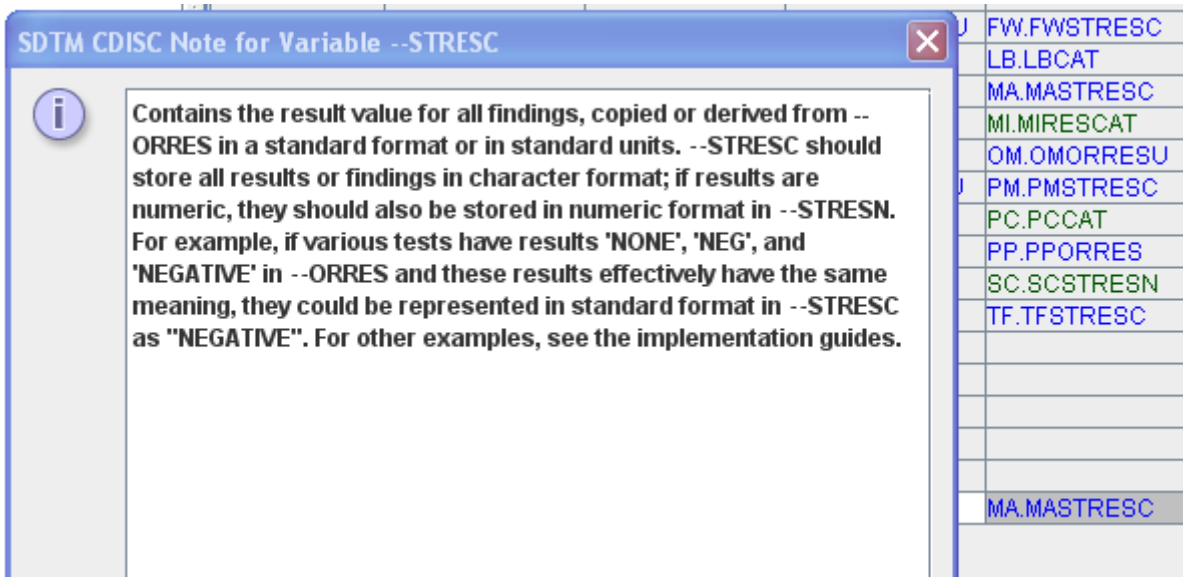
Implementation of SEND v.3.0 draft....3
Splitting of COVAL fields....4
Multiple COVAL fields when creating SQL to generate database tables....6
Multiple COVAL fields when cleaning up the define.xml file....7
Special versions for EDC vendor ODM-extensions....7

Implementation of SEND v.3.0 draft

The draft of the SEND standard v.3.0 has now been implemented.
When starting on a new set of mappings, the user has now the choice between SDTM v.1.1 (SDTM-IG 3.1.1), SDTM 1.2 (SDTM-IG 3.1.2) and SEND v.3.0 draft:



When SEND 3.0 is chosen, the define.xml template for SEND 3.0 is loaded, including all the necessary information from the SEND 3.0 draft standard, such as "CDISC notes":



Everything else works just like when working with one of the SDTM standards, except that the titles of some menus are being adapted.

Extended dialog and functionality for "Navigate -> Find hot SDTM candidate"

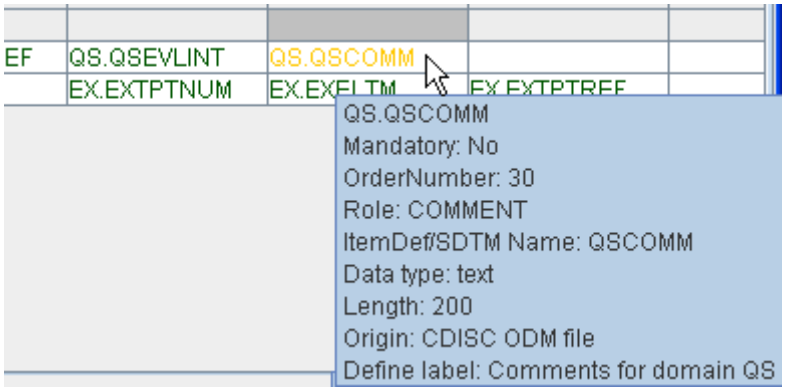
An extra field has been added to search in the ODM for items that are annotated with SDTM information using the "Alias" element. For example:

```
- <ItemDef DataType="integer" Length="3" Name="Systolic BP" OID="I_SYSBP" SASFieldName="SYSTL"
    SDSVarName="VSORRES">
  - <Question>
      <TranslatedText xml:lang="en">Systolic blood pressure</TranslatedText>
      <TranslatedText xml:lang="fr">Tension artérielle systolique</TranslatedText>
      <TranslatedText xml:lang="de">Systolischer Blutdruck</TranslatedText>
      <TranslatedText xml:lang="ko" />
    </Question>
    <MeasurementUnitRef MeasurementUnitOID="MU_MMHG" />
  + <RangeCheck Comparator="LT" SoftHard="Hard">
    <Alias Context="SDTM" Name="VSORRES where VSTESTCD=SYSBP" />
    <Alias Context="CDASH" Name="SYSBP" />
  </ItemDef>
```

When using the dialog "Find hot SDTM candidate", the ODM Item "I_SYSBP" will be found when the user searches for a suitable candidate for VSORRES or VSTESTCD, or when searching for "SYSBP":

Splitting of COVAL fields

Version 1.4 already had the possibility to add "Comment" variables to each SDTM domain, which are automatically transformed into CO-domain records at execution time.

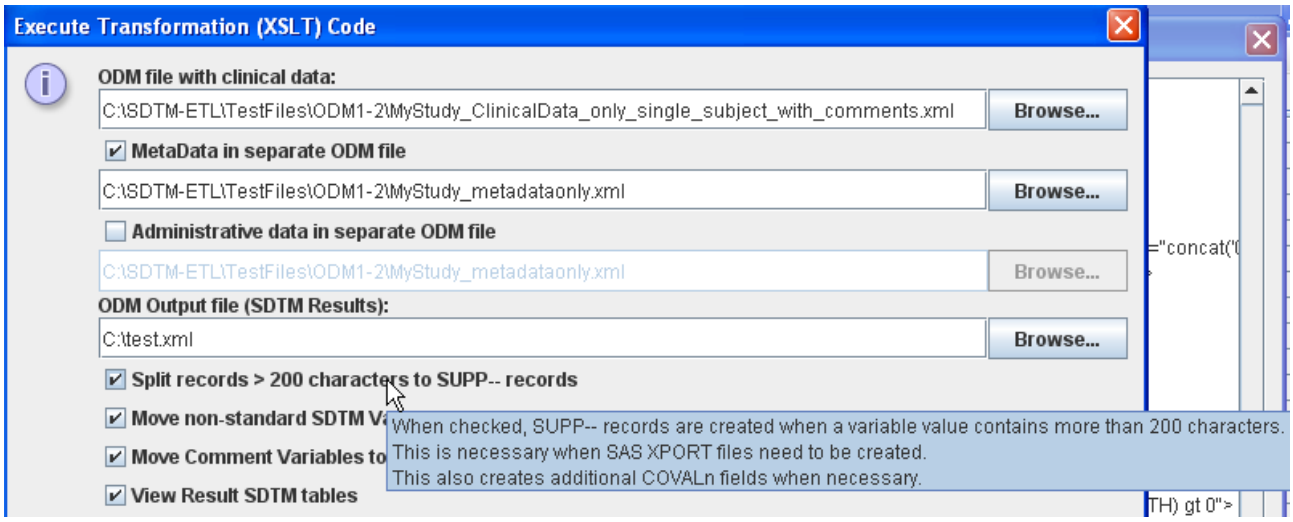


Together with the "comment()" function, that greps the content of the ODM "Annotation / Comment" element of an ItemData, this provides an efficient mechanism to generate CO-domain records automatically. The automatically generated CO-domain however only a single "COVAL" ("Comment Value") variable, which is just fine (and correct) when no SAS Transport 5 datasets need to be generated, and/or when all comments have less than 200 characters.

In case however that SAS Transport files need to be generated, and some comments have more than 200 characters, there is a risk of information loss.
Essentially, the 200-character limitation of SAS Transport 5 is a stupid one[1], and the "workaround" that was developed by the SDTM team is a pretty unfortunate one: when the contents of COVAL exceed 200 characters, additional fields (COVAL1, COVAL2, ...) each with a length of 200 characters need to be added to the CO-domain dataset.

Version 1.5 of the software now fully supports this: as well for automatically created CO-domains (from "Comment" variables in the normal SDTM datasets), as for explicit CO-domain instances, the software will perform a postprocessing step when requested by the user, and will for each of the CO-domain instances look for the length given for the COVAL variable, and create additional COVALn (i.e. COVAL1, COVAL2, ...) fields in the datasets when necessary, this as well in the SAS Transport 5 datasets, as in the XML datasets (although for the latter this isn't strictly necessary).

In order to allow for multiple COVAL fields, the user should check the checkbox "split records > 200 characters to SUPP-- records":

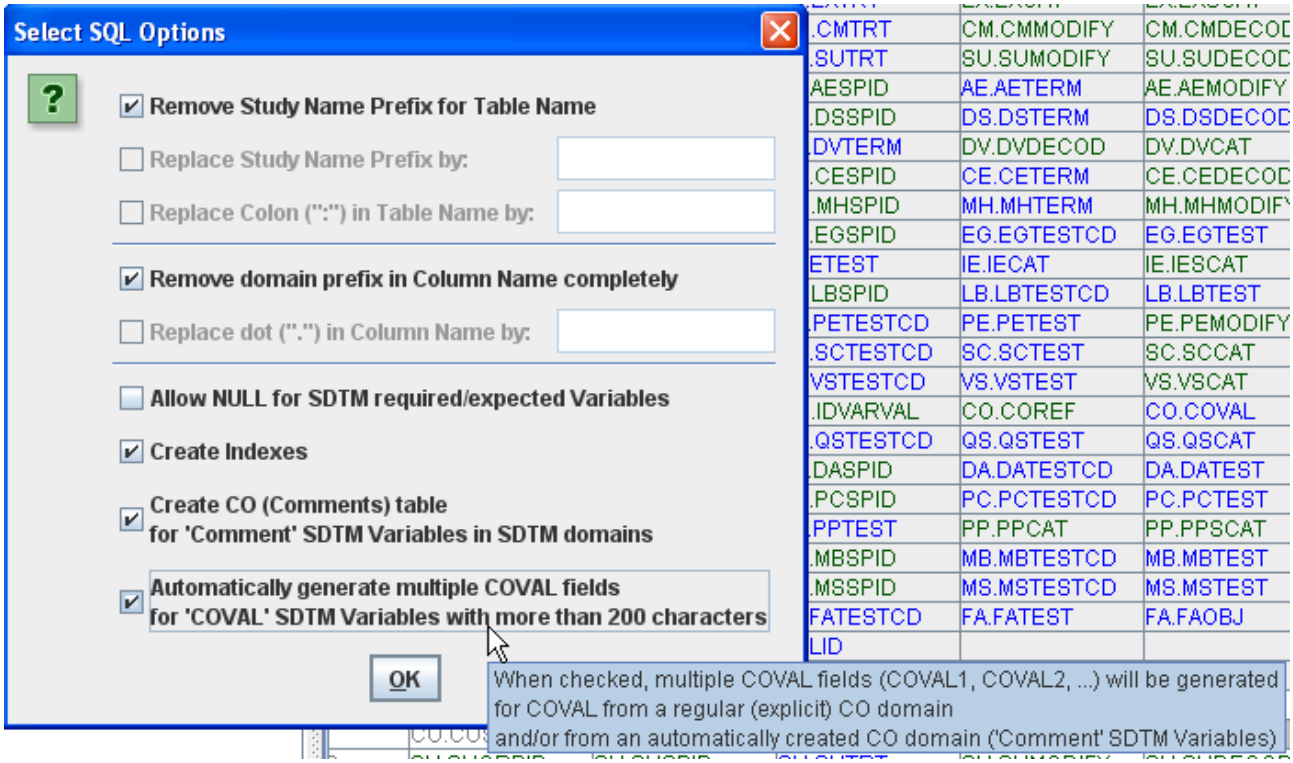


This will then also take care of automated generation of multiple COVAL fields when necessary.
The number of additional COVALn fields is determined by the given length of COVAL in the metadata. If the given length is 200 or less, no additional COVALn fields will be created.
If e.g. a length of 800 is given, four additional COVALn fields will be created (so five in total). This may seem too much, but one needs to take into account that splitting needs to be done between words, so that extra space may be needed.
For automatically created CO-domains from "Comment" variables in the regular datasets, the number of additional COVALn fields in the output is determined by the highest value for the length in the metadata of the "Comment" variables. So for example, if there is a "Comment" variable in the DM domain with length 100, and a "Comment" variable in the VS domain with length 700, then a CO domain will be created with COVAL, COVAL1, COVAL2 and COVAL3 fields.

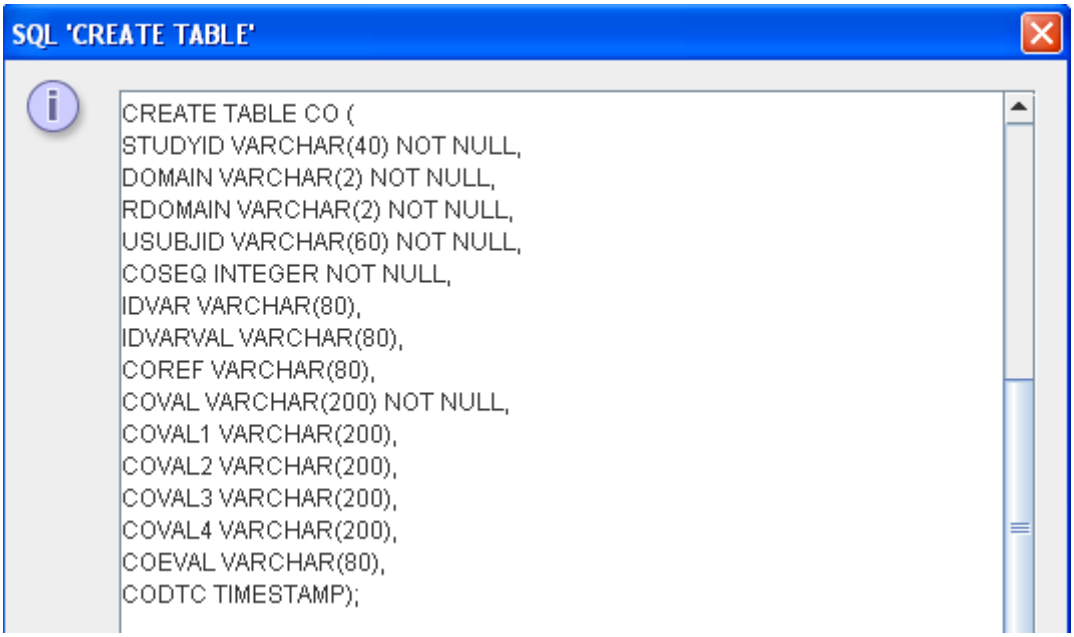
1 SAS Transport 5 is a very old computer format stemming from the age of the IBM mainframes, and is still only used by the FDA for electronic submissions. It needs urgent replacement, but the lack of knowledge of XML at the FDA has made this impossible sofar.

Multiple COVAL fields when creating SQL to generate database tables

Similarly, users may want to create SDTM databases with our without splitted COVAL fields for comments with more than 200 characters. In order to allow so, an extra item has been added to the dialog “Select SQL options”:

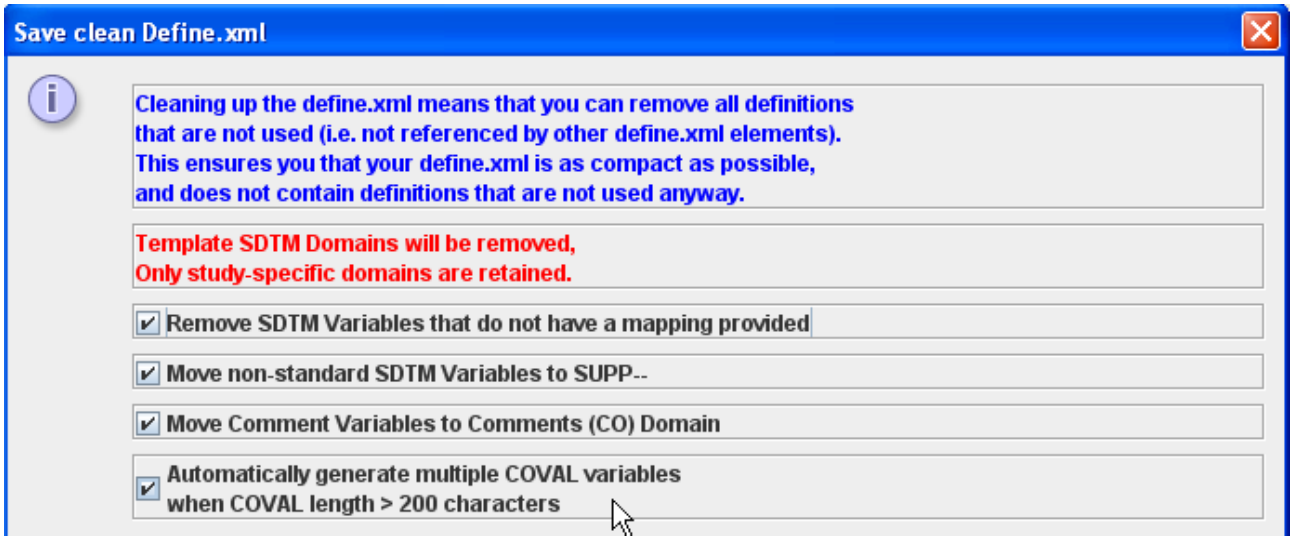


When checked, the number of additional COVALn fields in the “CREATE TABLE” statement that is generated will be determined by the length given for the COVAL field in the regular CO domain and/or the maximal length of any of the “Comment” SDTM variables in the regular domains.



Multiple COVAL fields when cleaning up the define.xml file

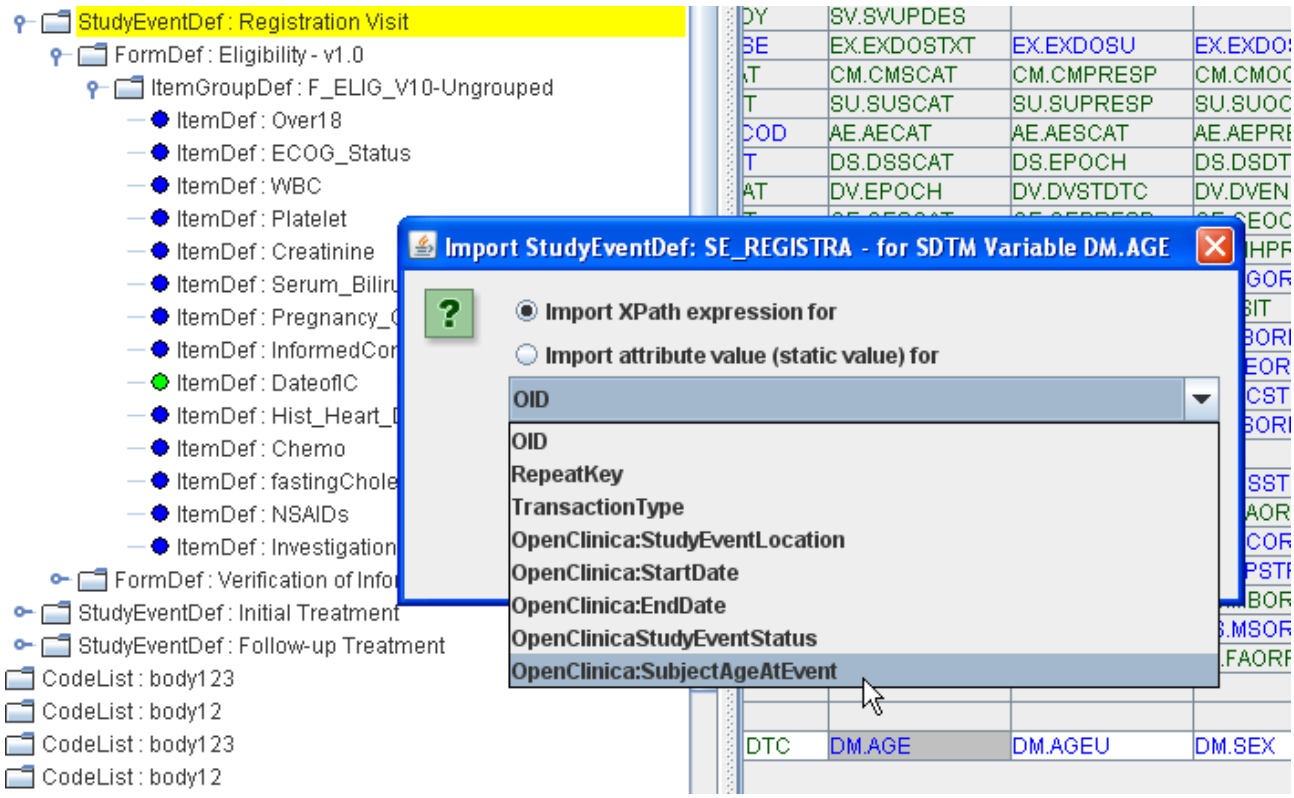
When using the dialog “File - Save cleaned define.xml”, the user will now also be asked whether he /she wants to have multiple COVAL fields in the exported define.xml.
It is adviced to do so when the cleaned define.xml file is meant for a submission to the FDA accompanied by SAS Transport 5 files, this as the latter do not support any fields longer than 200 characters, meaning that multiple COVAL fields need to be used when there are comments with more than 200 characters.
When doing so, the software will take care that the appropriate number of COVALn fields is generated in accordance with the rules given above.



Special versions for EDC vendor ODM-extensions

We are currently developing special versions of the software that have additional support for ODM-extensions of different EDC vendors.
The first special version that is now offered has support for the extensions of the ODM standard as has been implemented in OpenClinica 3.0.x.
This special version has been named “SDTM-ETL for OpenClinica Users”.
Other special versions are in preparation.

If you would like to have a special version for your specific EDC system, please contact us at info@XML4Pharma.com.


StudyEventDef : Registration Visit
FormDef : Eligibility - v1.0
ItemGroupDef : F_ELIG_V10-Ungrouped
ItemDef : Over18
ItemDef : ECOG_Status
ItemDef : WBC
ItemDef : Platelet
ItemDef : Creatinine
ItemDef : Serum_Biliru
ItemDef : Pregnancy_
ItemDef : InformedCor
ItemDef : DateofIC
ItemDef : Hist_Heart_
ItemDef : Chemo
ItemDef : fastingChole
ItemDef : NSAIDs
ItemDef : Investigation
FormDef : Verification of Infor
StudyEventDef : Initial Treatment
StudyEventDef : Follow-up Treatment
CodeList : body123
CodeList : body12
CodeList : body123
CodeList : body12
Import StudyEventDef: SE_REGISTRA - for SDTM Variable DM.AGE
Import XPath expression for
Import attribute value (static value) for
OID
OID
RepeatKey
TransactionType
OpenClinica:StudyEventLocation
OpenClinica:StartDate
OpenClinica:EndDate
OpenClinicaStudyEventStatus
OpenClinica:SubjectAgeAtEvent
DY
SV.SVUPDES
EX.EXDOSTXT
EX.EXDOSU
CM.CMSCAT
CM.CMPRESP
SU.SUSCAT
SU.SUPRESP
AE.AECAT
AE.AESCAT
DS.DSSCAT
DS.EPOCH
DV.EPOCH
DV.DVSTDTC
DTC
DM.AGE
DM.AGEU
DM.SEX