**Step-by-step description of the algorithm to establish continuous percentile charts**

1. Partition of input data into overlapping age groups

Samples are split into age-specific groups (bins) containing 5,000 samples (or 2,500 samples for children < 100 days).

*Details:* Each bin is filled until it contains 5,000/2,500 samples, this is achieved by widening the age interval for each bin until 5,000/2,500 samples from different patients are included. In samples from children < 100 days, if multiple samples from an individual patient are contained in the age interval, only the sample nearest to the age interval center is selected. The split process is started at *age = 0 days*, the first bin is then filled; whenever a bin contains *≥ ½ × n* samples during the filling process, the upper limit of that age interval is selected as the center of the age interval for the subsequent bin (to create overlapping age groups). This is repeated until *age > 18 years*.

*Special case:* For some age groups (e.g. *age = 0 days*, *age = 1 days*), more than 5,000/2,500 samples are available. In that case, the bin contains all samples with the given age and the subsequent day is selected as the center of the subsequent age interval. (E.g. *bin0 days* and *bin1 day* contain all samples from different patients with *age = 0 days* and *age = 1 day*, respectively, while *bin50 days* contains samples from different patients from an age interval centered on *age = 50 days*.)

2. Generation of discrete reference intervals for each age group via an indirect method

For each age-specific group (bin) established in step 1, the distribution of physiological samples is calculated using the *Reference Limit Estimator* developed by the *DGKL Working Group for Guide Limits*, available at [http://www.dgkl.de/](http://www.dgkl.de/PA106975_DE_VAR100), with setting *Pathological values = “Both”*.

3. Conversion of discrete reference intervals into continuous percentile charts

All 2.5th, 50th and 97.5th percentiles resulting from step 3 are converted to continuous intervals using the R software's smooth.spline function. The number of degrees of freedom is chosen by visual inspection of the curves generated to account for the imbalance of available data points and physiological dynamics (automated techniques for the selection of degrees of freedom favor quantity of data points for curve fitting [i.e. number of age groups] over dynamics represented by fewer data points [e.g. changes in the neonatal period]). The mentioned imbalance also results in the creation of two different curves to adequately represent the different dynamics of the neonatal period and infancy on the one hand and toddlerhood to adolescence on the other hand.

**Comparison to our previous approach**

To enable comparison of the results established in the current study using the enhanced approach (i.e. with extensive filtering of samples) to the approach used in our previous publications (i.e. without filtering of samples), we additionally applied our previous approach to the current dataset. Specifically, we included all samples (irrespectively of retesting) in the analysis and used a constant age bin size n=5,000 for all age ranges (as described above and implemented in our previous analyses, only one sample per patient is included in each age bin).

**Analysis of influencing factors**

To evaluate the influence of samples where test results in other analytes were outside the generated reference limits, we performed a sub-analysis excluding these samples. We excluded all samples with test results below the 2.5th or above the 97.5th percentile for hemoglobin, hematocrit, red cell count, white cell count, platelet count, or RDW (except for the examined analyte), for RDW analysis we excluded samples with abnormal results for hemoglobin, hematocrit, white cell count, red cell count, or platelet count, and for the analysis of red cell indices we excluded samples with abnormal results for white cell count, platelet count, RDW, and hemoglobin (for MCV), hematocrit (for MCH), or red cell count (for MCHC).

Differences due to family origin were explored in a similar sub-analysis, in which we excluded children of Turkish descent (the major ethnic minority in Germany (13)), as these children may display variation in hematopoiesis due to Turkey’s Mediterranean location. Children were identified using a variant of a validated algorithm which evaluates a person’s name to assess Turkish descent (21,22). We adapted this otherwise highly specific algorithm to a more sensitive (and less specific) variant, such that all children with family names and first names mainly of Turkish (and possibly other Middle Eastern) descent were excluded from this sub-analysis (6.2% of patients and 6.9% of samples were excluded). This analysis was performed for all children from center A, as patient names from other centers were anonymized before transfer to the main study center to comply with German privacy laws. To account for the reduced sample size (center A only), and as this sub-analysis focused on differences between groups rather than absolute reference intervals, we did not perform the filtering step described above for this sub-analysis.

**Comparison with existing reference intervals**

We compared our results to percentile charts from the German KiGGS study (23). In that study, blood counts were performed on samples from 14,076 healthy children recruited from the community (aged 1–17 years) using an Abbott Cell-Dyn 3500 hematology analyzer. Results for hemoglobin, hematocrit, red cell count, and red cell indices have been published (3rd, 5th, 10th, 25th, 50th, 75th, 90th, 95th, and 97th percentiles for every ½ year from 1½ years to 17½ years); the mathematical descriptions of the percentile charts and white cell count and platelet data are not available. To enable comparison of our results to the KiGGS data, we recalculated the parameters of the LMS distribution (λ, µ, σ) underlying the published tables and computed the corresponding 2.5th and 97.5th percentiles.

**Calculation of z-scores**

Z-Scores were calculated according to the following formula (i.e. z-normalization after Box-Cox transformation of values):

*yλ(age)*, *μ(age)*, and *σ(age)* are the age-specific parameters of the distribution of samples underlying the percentile charts.

This normalization results in z-scores between -2 and +2 for values considered normal (-2 corresponds to the 2.3rd percentile and +2 corresponds to the 97.7th percentile) and a z-score of ±0 for the median test result.

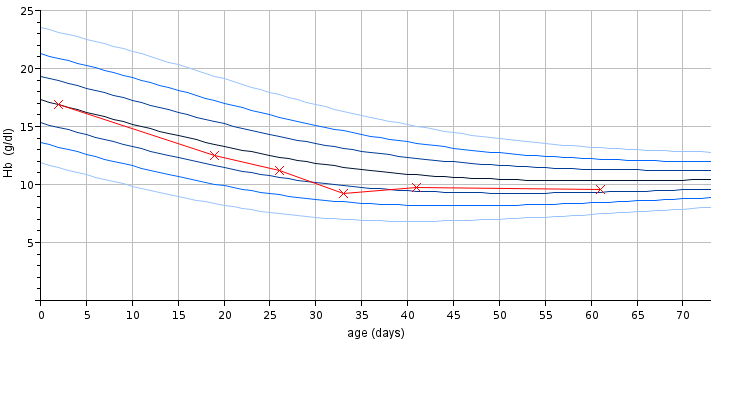
While normalization using z-scores is recognized internationally, other approaches exist and may be considered more suitable for human interpretation. E.g. “quantity quotient reporting” uses 100 as the median and 80 to 120 as the range of normal (95 % reference interval) (12); conversion of z-scores to quantity quotients is possible.

**Presentation for the evaluation of percentile charts in comparison to conventional representations of laboratory test results**

The following eight clinical scenarios were shown to physicians during training meetings using a PowerPoint® (Microsoft, United States) presentation and each participant’s decision times and answer correctness were measured using single-choice remote controls (PowerVote®, La Générale Multimédia, France). Each scenario and its associated laboratory test results had to be evaluated using both graphical and conventional representations in alternating order, and clinicians were told that the laboratory test results for each representation might or might not be the same to enforce independent evaluation of each representation. For both evaluations, preliminary reference intervals and percentile charts from this report were shown, i.e. only the presentation of test results and reference intervals/percentiles differed while the underlying test results were identical. Correct answers were defined beforehand and in some scenarios multiple answers are considered correct (these scenarios focus on the physicians’ time to assess the hematology dynamics). These tests were performed at three German pediatric tertiary care centers (n = 122 pediatricians) in Erlangen, Augsburg, and Dresden (slides and presentations were in German, slides in Dresden were modified to show mmol/l for hemoglobin). Due to time constraints, scenario 4 and 7 were only shown on one occasion in Erlangen (and not in Augsburg and Dresden).

**Scenario 1**

* A newborn of an HIV-positive mother receives Retrovir® for the prevention of mother-to-child-transmission of HIV.
* Retrovir® can cause anemia.
* Blood counts are performed at regular check-ups.
* **Evaluate the course of hemoglobin over time. Is there a drop in hemoglobin over time? If so, is it clinically relevant?**
* Answers
  1. Unremarkable course of hemoglobin over time. (Correct answer.)
  2. Drop in hemoglobin, no need for intervention.
  3. Drop in hemoglobin, intervention needed.
* First representation (percentile charts):



* Second representation (conventional/tabular representation):

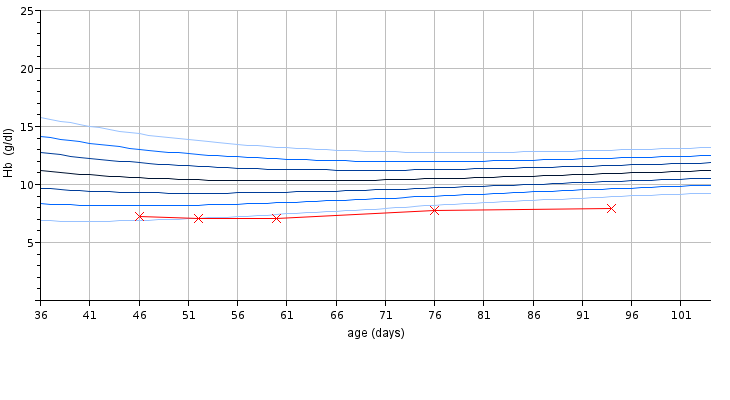
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Age** | **2 days** | **19 days** | **26 days** | **33 days** | **41 days** | **61 days** |
| **Hb (g/dl)** | 16.9 | 12.5 | 11.2 | 9.2 | 9.7 | 9.6 |
| **RI** | 11.4–23.1 | 8.4–19.3 | 7.5–17.7 | 7.0–16.3 | 6.8–15.0 | 7.5–13.2 |

**Scenario 2**

* An infant with hereditary spherocytosis receives regular check-ups, including blood counts.
* **Evaluate the course of hemoglobin over time and assess the course of anemia in this child.**
* Answers
  1. Increasing anemia. (Correct answer.)
  2. Stable anemia.
  3. Decreasing anemia.
* First representation (conventional/tabular representation):

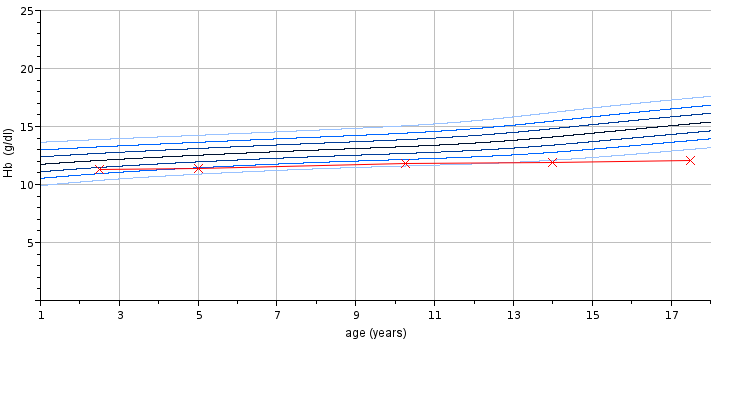
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Age** | **46 days** | **52 days** | **60 days** | **76 days** | **94 days** | **46 days** |
| **Hb (g/dl)** | 7.2 | 7.1 | **7.1** | **7.8** | **7.9** | 7.2 |
| **RI** | 6.9–14.4 | 7.1–13.8 | 7.4–13.2 | 8.2-12.8 | 8.9–13.0 | 6.9–14.4 |

* Second representation (percentile charts):



**Scenario 3**

* Regular check-ups are perfomed in a boy (not due to a primarily hematological disease).
* **Evaluate the course of hemoglobin over time.**
* Answers
  1. Rise in hemoglobin.
  2. Unremarkable course of hemoglobin over time.
  3. Drop in hemoglobin. (Correct answer.)
* First representation (percentile charts):

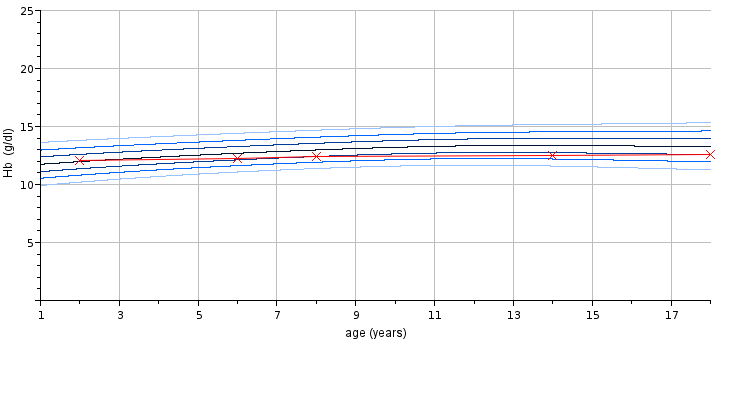


* Second representation (conventional/tabular representation):

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Age** | **2J 6M** | **5J** | **10J 3M** | **14J** | **17J 6M** |
| **Hb (g/dl)** | 11.3 | 11.4 | 11.8 | **11.9** | **12.1** |
| **RI** | 10.4–13.9 | 10.9-14.2 | 11.6-15.1 | 12.1–16.2 | 13.0–17.4 |

**Scenario 4**

* Regular check-ups are perfomed in a girl (not due to a primarily hematological disease).
* **Evaluate the course of hemoglobin over time.**
* Answers
  1. Rise in hemoglobin.
  2. Unremarkable course of hemoglobin over time. (Correct answer [both 2 and 3].)
  3. Drop in hemoglobin. (Correct answer [both 2 and 3].)
* First representation (percentile charts):

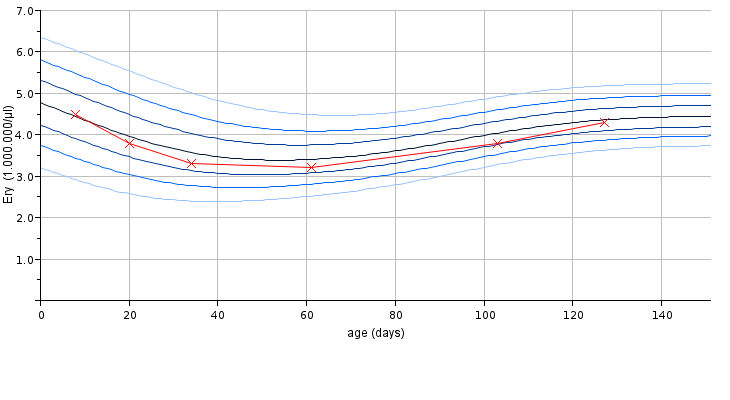


* Second representation (conventional/tabular representation):

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Age** | **2 years** | **6 years** | **8 years** | **14 years** | **18 years** |
| **Hb (g/dl)** | 12.1 | 12.2 | 12.4 | 12.5 | 12.6 |
| **RI** | 10.2–13.8 | 11.1–14.4 | 11.4–14.7 | 11.6–15.2 | 11.3–15.3 |

**Scenario 5**

* **Evaluate the course of red cell count over time.**
* Answers
  1. The drop in red cell count is physiological. (Correct answer.)
  2. The drop in red cell count is pathological.
* First representation (percentile charts):



* Second representation (conventional/tabular representation):

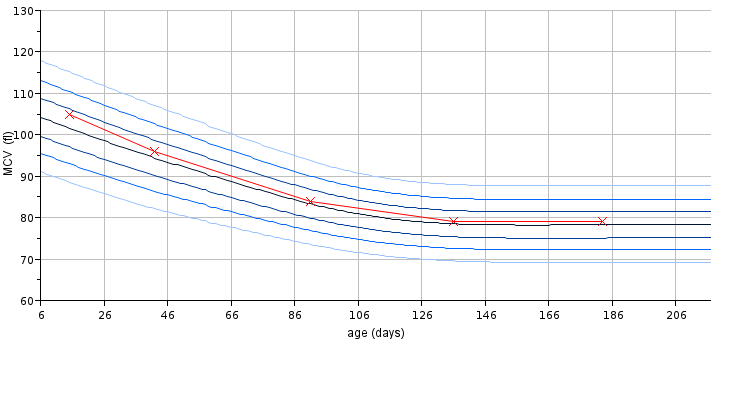
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Age** | **8 days** | **20 days** | **34 days** | **2 months** | **3 months 12 days** | **4 months 5 days** |
| **Red cell count (109/µl)** | 4.5 | 3.8 | 3.3 | 3.2 | 3.8 | 4.3 |
| **RI** | 2.9–6.0 | 2.6–5.5 | 2.4–5.0 | 2.5–4.5 | 3.3–4.9 | 3.6–5.2 |

**Scenario 6**

* Regular measurements of MCV are performed in a newborn/infant boy.
* **Evaluate the course of MCV over time.**
* Answers
  1. Microcytosis occurs.
  2. Physiological course of MCV over time. (Correct answer.)
* First representation (conventional/tabular representation):

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Age** | **15 days** | **1 months 12 days** | **3 months** | **4 months 14 days** | **6 months** |
| **MCV (fl)** | 105 | 96 | 84 | 79 | 79 |
| **RI** | 88.6–115.2 | 82.2–107.1 | 73.5–93.7 | 69.6–88.0 | 69.2–87.8 |

* Second representation (percentile charts):

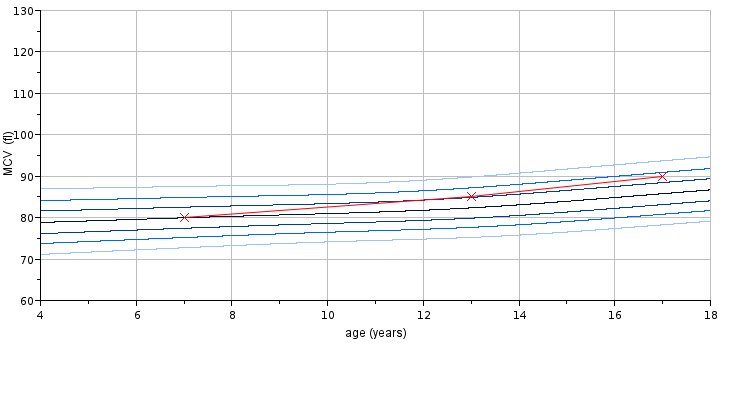


**Scenario 7**

* **Evaluate the course of MCV over time.**
* Answers
  1. Physiological course of MCV over time. (Correct answer [both 1 and 2].)
  2. The rise in MCV over time is pathological, e.g. due to vitamin B12 deficiency. (Correct answer [both 1 and 2].)
* First representation (conventional/tabular representation):

|  |  |  |  |
| --- | --- | --- | --- |
| **Age** | **7 years** | **13 years** | **17 years** |
| **MCV (fl)** | 80 | 85 | 90 |
| **RI** | 72.7–87.5 | 75.2–89.8 | 78.3–93.7 |

* Second representation (percentile charts):

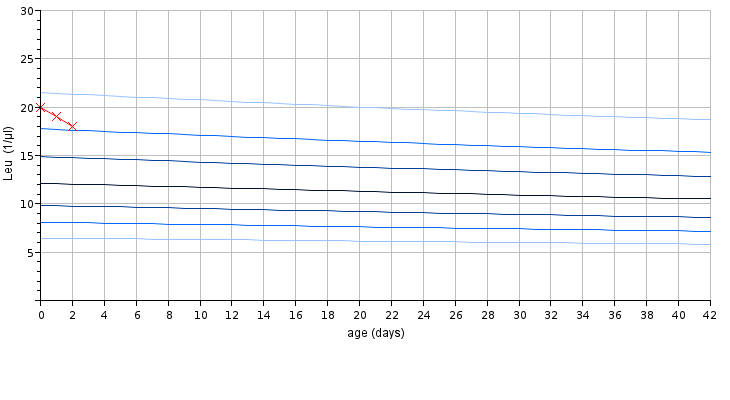


**Scenario 8**

* **Evaluate the white cell counts in a neonate.**
* Answers
  1. The measurements are physiological. (Correct answer.)
  2. The measurements are pathological and might indicate infection.
* First representation (conventional/tabular representation):

|  |  |  |  |
| --- | --- | --- | --- |
| **Age** | **0 days** | **1 day** | **2 days** |
| **WBC (10³/µl)** | 20.0 | 19.0 | 18.0 |
| **RI** | 6.5–21.5 | 6.4–21.4 | 6.4–21.4 |

* Second representation (percentile charts):



**Evaluation of z-scores in comparison to absolute test results for the discrimination of myelodysplastic syndrome and severe aplastic anemia using logistic regression and support vector machines**

We evaluated blood counts from children with myelodysplastic syndrome (MDS) and severe aplastic anemia (SAA) from the EWOG-MDS and EWOG-SAA trial databases. The first available blood count (hemoglobin, red cell count, mean cell volume [MCV], platelet count, and white cell count) prior to any transfusions was selected (n = 379 [MDS], including n = 225 children with refractory cytopenia of childhood [RCC], and n = 39 [SAA]). Using logistic regression and support vector machines we evaluated the power of absolute test results in comparison to z-scores to discriminate between MDS and SAA and MDS-RCC and SAA. Implemented in custom Python scripts and scikit-learn (17) we trained the models using all but one randomly selected dataset from either group (MDS/MDS-RCC and SAA) and predicted the remaining two datasets (MDS/MDS-RCC and SAA). This was repeated *n\_iter = 100 × n\_bootstrap = 100* times, resulting in 10,000 predictions for both z-scores and absolute test results (a new model was generated for each prediction and no data from the former models was used).

The essential Python code fragment is shown below.

**import** pandas **as** pd  
**import** numpy **as** np  
**from** sklearn **import** preprocessing  
**from** sklearn.svm **import** SVC  
**from** sklearn.linear\_model **import** LogisticRegression  
  
data = pd.read\_csv(**"EWOG-DATA.csv"**)  
data = data[(data[**"THR"**].notnull())& (data[**"MCV"**].notnull()) & (data[**"HB"**].notnull())& (data[**"ERY"**].notnull()) & (data[**"LEU"**].notnull())]  
  
kernel = **"sigmoid"**n\_test = 1  
n\_bootstrap = 100  
n\_iter = 100  
  
mds = data[data[**"MDS/SAA"**] == **"MDS"**]  
saa = data[data[**"MDS/SAA"**] == **"SAA"**]  
  
sums\_svc\_std = []  
sums\_svc\_z = []  
sums\_logreg\_std = []  
sums\_logreg\_z = []  
  
**for** \_ **in** range(n\_bootstrap):  
   sum\_svc\_std = 0  
   sum\_svc\_z = 0  
   sum\_logreg\_std = 0  
   sum\_logreg\_z = 0  
  
   **for** \_ **in** range(n\_iter):  
x = np.array([np.concatenate((mds[**"THR"**].values, saa[**"THR"**].values)),  
         np.concatenate((mds[**"MCV"**].values, saa[**"MCV"**].values)),  
         np.concatenate((mds[**"HB"**].values, saa[**"HB"**].values)),  
         np.concatenate((mds[**"ERY"**].values, saa[**"ERY"**].values)),  
         np.concatenate((mds[**"LEU"**].values, saa[**"LEU"**].values))]).transpose()  
      x = preprocessing.scale(x)  
      xz = np.array([  
np.concatenate((mds[**"THR-Z"**].values, saa[**"THR-Z"**].values)),  
         np.concatenate((mds[**"MCV-Z"**].values, saa[**"MCV-Z"**].values)),  
         np.concatenate((mds[**"HB-Z"**].values, saa[**"HB-Z"**].values)),  
         np.concatenate((mds[**"ERY-Z"**].values, saa[**"ERY-Z"**].values)),  
         np.concatenate((mds[**"LEU-Z"**].values, saa[**"LEU-Z"**].values))  
]).transpose()  
      xz = preprocessing.scale(xz)  
      y = np.array([0] \* len(mds) + [1] \* len(saa))  
  
      test\_indices\_mds = np.random.choice(range(len(mds)), n\_test)  
      test\_indices\_saa = np.random.choice(range(len(mds), len(mds)+len(saa)), n\_test)  
  
train\_indices\_mds = [i **for** i **in** range(len(mds)) **if** i **not in** test\_indices\_mds]  
      train\_indices\_saa = [i **for** i **in** range(len(mds), len(mds)+len(saa)) **if** i **not in** test\_indices\_saa]  
  
      svc = SVC(kernel=kernel)  
      svc.fit(x[np.concatenate([train\_indices\_mds, train\_indices\_saa])], y[np.concatenate([train\_indices\_mds, train\_indices\_saa])])  
      sum\_svc\_std += np.sum(svc.predict(x[np.concatenate([test\_indices\_mds, test\_indices\_saa])]) == y[np.concatenate([test\_indices\_mds, test\_indices\_saa])])  
  
      svcz = SVC(kernel=kernel)  
      svcz.fit(xz[np.concatenate([train\_indices\_mds, train\_indices\_saa])], y[np.concatenate([train\_indices\_mds, train\_indices\_saa])])  
      sum\_svc\_z += np.sum(svcz.predict(xz[np.concatenate([test\_indices\_mds, test\_indices\_saa])]) == y[np.concatenate([test\_indices\_mds, test\_indices\_saa])])  
  
      lr = LogisticRegression()  
      lr.fit(x[np.concatenate([train\_indices\_mds, train\_indices\_saa])], y[np.concatenate([train\_indices\_mds, train\_indices\_saa])])  
      sum\_logreg\_std += np.sum(lr.predict(x[np.concatenate([test\_indices\_mds, test\_indices\_saa])]) == y[np.concatenate([test\_indices\_mds, test\_indices\_saa])])  
  
      lrz = LogisticRegression()  
      lrz.fit(xz[np.concatenate([train\_indices\_mds, train\_indices\_saa])], y[np.concatenate([train\_indices\_mds, train\_indices\_saa])])  
      sum\_logreg\_z += np.sum(lrz.predict(xz[np.concatenate([test\_indices\_mds, test\_indices\_saa])]) == y[np.concatenate([test\_indices\_mds, test\_indices\_saa])])  
  
   sums\_svc\_std += [sum\_svc\_std]  
   sums\_svc\_z += [sum\_svc\_z]  
   sums\_logreg\_std += [sum\_logreg\_std]  
   sums\_logreg\_z += [sum\_logreg\_z]