Validity of a simplified risk score model

A simple clinical risk score model for predicting child malnutrition in a limited resource setting such as Kenya could be valuable in several ways. Such a model could help to identify children at risk of malnutrition and enable early intervention to prevent the development of malnutrition and its associated complications. This could improve the health outcomes of children and reduce the burden on healthcare systems.

[One example of such a model is the Screening Tool for Risk of Impaired Nutritional Status and Growth (STRONGkids), which has been used to estimate the risk of malnutrition in hospitalized children1](https://ijponline.biomedcentral.com/articles/10.1186/1824-7288-39-81). [The STRONGkids consists of 4 items providing a score that classifies a patient in low, moderate, high risk for malnutrition1](https://ijponline.biomedcentral.com/articles/10.1186/1824-7288-39-81). [According to STRONGkids, children at high risk for malnutrition had significantly lower Height for Age values and BMI values in comparison to other groups1](https://ijponline.biomedcentral.com/articles/10.1186/1824-7288-39-81).

Current trends and patterns in child malnutrition indicate that this is a significant global issue. [In 2020, globally, 149.2 million children under the age of 5 years were stunted, 45.4 million wasted, and 38.9 million overweight2](https://www.who.int/publications-detail-redirect/9789240025257). [The number of children with stunting is declining in all regions except Africa2](https://www.who.int/publications-detail-redirect/9789240025257). These figures highlight the need for effective tools to identify and address malnutrition in children.

In summary, a simple clinical risk score model for predicting child malnutrition could be justified by its potential to improve health outcomes for children and reduce the burden on healthcare systems. The current trends and patterns in child malnutrition indicate that this is a significant global issue that requires effective tools to address.

1 [Application of a score system to evaluate the risk of malnutrition in a multiple hospital setting | Italian Journal of Pediatrics | Full Text (biomedcentral.com)](https://ijponline.biomedcentral.com/articles/10.1186/1824-7288-39-81)

2 [Levels and trends in child malnutrition: UNICEF/WHO/The World Bank Group joint child malnutrition estimates: key findings of the 2021 edition](https://www.who.int/publications/i/item/9789240025257)

3 [Management of moderate acute malnutrition in children in resource-limited settings - UpToDate](https://www.uptodate.com/contents/management-of-moderate-acute-malnutrition-in-children-in-resource-limited-settings)

4 [A four-stage evaluation of the Paediatric Yorkhill Malnutrition Score in a tertiary paediatric hospital and a district general hospital | British Journal of Nutrition | Cambridge Core](https://www.cambridge.org/core/journals/british-journal-of-nutrition/article/fourstage-evaluation-of-the-paediatric-yorkhill-malnutrition-score-in-a-tertiary-paediatric-hospital-and-a-district-general-hospital/2F57A5711A1D5EB7E33548FE677402E2)

5 [A four-stage evaluation of the Paediatric Yorkhill Malnutrition Score in a tertiary paediatric hospital and a district general hospital | British Journal of Nutrition | Cambridge Core](https://www.cambridge.org/core/journals/british-journal-of-nutrition/article/fourstage-evaluation-of-the-paediatric-yorkhill-malnutrition-score-in-a-tertiary-paediatric-hospital-and-a-district-general-hospital/2F57A5711A1D5EB7E33548FE677402E2)

* Fit a lasso model to predict stunting, underweight, and wasting using all covariates (or use only the sig vars??)
* Evaluate the performance of selected variables from the lasso reg model in predicting each of stun, wast and und
* Compute the specificity sensitivity and AUC for each variable returned by the lasso reg model.

To prove that \(E[Y\_a] = E[E[Y|X, A=a, R=1]]\) given that \(Y \perp R\) (i.e., \(Y\) is independent of \(R\)), we will use the law of total expectation. This law states that for any random variables \(X\) and \(Y\), and any event \(A\), we have \(E[Y] = E[E[Y|X]]\). We will apply this law to the given expression.

We're given the following variables:

- \(Y\): Observed outcome.

- \(R\): Missingness indicator (1 if \(Y\) is observed, 0 if missing).

- \(Y\_a\): Potential outcome.

- \(X\): Covariates.

- \(A\): Treatment.

We want to prove that \(E[Y\_a] = E[E[Y|X, A=a, R=1]]\).

First, let's start with the right-hand side of the expression:

\[E[E[Y|X, A=a, R=1]]\]

By the law of total expectation, we can rewrite the inner expectation as follows:

\[E[Y|X, A=a, R=1] = E[Y|X, A=a, R=1, Y=1] \cdot P(Y=1|X, A=a, R=1) + E[Y|X, A=a, R=1, Y=0] \cdot P(Y=0|X, A=a, R=1)\]

Since \(Y\) is independent of \(R\) (given in the problem statement), we can factor out \(R=1\) and write:

\[E[Y|X, A=a, R=1] = E[Y|X, A=a, Y=1] \cdot P(Y=1|X, A=a, R=1) + E[Y|X, A=a, Y=0] \cdot P(Y=0|X, A=a, R=1)\]

Now, consider that when \(Y=0\), \(Y\_a\) is the potential outcome that corresponds to \(Y\) being equal to 0. Similarly, when \(Y=1\), \(Y\_a\) corresponds to the potential outcome for \(Y\) being equal to 1. Therefore, we can write:

\[E[Y|X, A=a, Y=0] = E[Y\_a|X, A=a, Y=0] \quad \text{and} \quad E[Y|X, A=a, Y=1] = E[Y\_a|X, A=a, Y=1]\]

Substituting these expressions back into our previous equation:

\[E[Y|X, A=a, R=1] = E[Y\_a|X, A=a, Y=1] \cdot P(Y=1|X, A=a, R=1) + E[Y\_a|X, A=a, Y=0] \cdot P(Y=0|X, A=a, R=1)\]

Now, we take the outer expectation with respect to \(E[Y\_a]\):

\[E[E[Y|X, A=a, R=1]] = E[E[Y\_a|X, A=a, Y=1] \cdot P(Y=1|X, A=a, R=1) + E[Y\_a|X, A=a, Y=0] \cdot P(Y=0|X, A=a, R=1)]\]

Using linearity of expectation, we can split this into two terms:

\[E[E[Y\_a|X, A=a, Y=1] \cdot P(Y=1|X, A=a, R=1)] + E[E[Y\_a|X, A=a, Y=0] \cdot P(Y=0|X, A=a, R=1)]\]

For each term, we can factor out the part that doesn't depend on \(Y\) since \(Y\) and \(Y\_a\) are independent:

\[E[Y\_a|X, A=a, Y=1] \cdot E[P(Y=1|X, A=a, R=1)] + E[Y\_a|X, A=a, Y=0] \cdot E[P(Y=0|X, A=a, R=1)]\]

Since \(Y\) and \(R\) are independent, \(P(Y=1|X, A=a, R=1) = P(Y=1|X, A=a)\) and \(P(Y=0|X, A=a, R=1) = P(Y=0|X, A=a)\), and we have:

\[E[E[Y\_a|X, A=a, Y=1] \cdot E[P(Y=1|X, A=a)]] + E[E[Y\_a|X, A=a, Y=0] \cdot E[P(Y=0|X, A=a)]]\]

Since \(Y\) and \(R\) are independent, we can also say that \(Y\) and \(A\) are independent, which means \(P(Y|X, A=a) = P(Y|A=a)\):

\[E[E[Y\_a|X, A=a, Y=1] \cdot E[P(Y=1|A=a)]] + E[E[Y\_a|X, A=a, Y=0] \cdot E[P(Y=0|A=a)]]\]

Now, notice that \(E[Y\_a|X, A=a, Y=1]\) and \(E[Y\_a|X, A=a, Y=0]\) are just constants with respect to the outer expectation, since they don't depend on the inner expectation variables. We can factor them out:

\[E[Y\_a|X, A=a, Y=1] \cdot E[P(Y=1|A=a)] + E[Y\_a|X, A=a, Y=0] \cdot E[P(Y=0|A=a)]\]

Finally, since \(Y\_a\) is a potential outcome and does not depend on the observed values, we can treat it as a constant when taking the expectation:

\[E[Y\_a|X, A=a, Y=1] \cdot P(Y=1|A=a) + E[Y\_a|X, A=a, Y=0] \cdot P(Y=0|A=a)\]

This expression is equivalent to \(E[Y\_a]\), the expected potential outcome, which completes the proof:

\[E[E[Y|X, A=a, R=1]] = E[Y\_a]\]

Therefore, we have proven that \(E[Y\_a] = E[E[Y|X, A=a, R=1]]\) given that \(Y \perp R\).