NeoPrediX
Forecast, save and improve lives

# Evaluation of NeoPrediX B.1 in a prospective observational multi-center study

NeoPrediX AG

## BACKGROUND

Neonatal jaundice is common and mostly transient without the need for intervention. To identify newborn infants at risk of bilirubin toxicity, bilirubin is measured several times after birth and before discharge. As most infants are discharged two days after birth but peak bilirubin concentrations are reached a couple of days later, parents are often asked to return with their newborn baby to the hospital one or two days after discharge. Bilirubin is then checked again to prevent that any infants requiring treatment due to hyperbilirubinemia are missed. A precise prediction of bilirubin progression over time, based on measurements taken during the initial hospital stay after birth, would be desirable to avoid unnecessary follow-ups and rehospitalizations. This prospective clinical multi-center study evaluates the recently developed NeoPrediX B.1 algorithm, which predicts bilirubin progression up to 48 hours into the future based on a few bilirubin measurements and clinical variables.

## METHODS AND NEONATAL STUDY POPULATION

Newborn infants were recruited between August 9th, 2021, and November 29th, 2021, at two university hospitals in Germany. The study protocol defined the following inclusion criteria: gestational age at birth ≥ 34+0/7 weeks and first bilirubin measurement earlier than 72 hours after birth. The exclusion criteria were defined as: birth weight < 1500 g, phototherapy or exchange transfusion prior to inclusion, genetically defined syndrome, and severe congenital malformation adversely affecting life expectancy or admission for a priori planned palliative care.

After applying inclusion and exclusion criteria, a total of n = 276 newborn infants with suitable bilirubin measurements were included into the study for analysis. Both, total serum bilirubin (TSB) as well as transcutaneous bilirubin (TcB) measured with the Jaundice Meter JM-105, were determined and five different scenarios were explored. The five scenarios differed in prediction horizon (up to 24 or 48 hours) and number of TSB and/or TcB measurements applied to the NeoPrediX B.1 algorithm.

**Scenario 1 with one TSB measurement (up to 24 h)**
At least one TSB measurement between 8 and 72 hours of postnatal age is needed for prediction up to 24 hours. The algorithm makes a prediction based on exactly one TSB measurement.

**Scenario 2 with TSB measurements (up to 48 h)**
At least two measurements are needed for prediction up to 48 hours. The algorithm requires the initial TSB to be between 8 and 72 hours of postnatal age and at least one additional TSB between 24 and 96 hours of postnatal age, with ≥ 8 hours between two consecutive measurements.

**Scenario 3 with TSB and TcB measurements (up to 48 h)**
At least two measurements are needed for prediction up to 48 hours. The algorithm requires the initial TSB measurement to be between 8 and 72 hours of postnatal age and at least one subsequent TcB measurement between 24 and 96 hours of postnatal age, with ≥ 8 hours between two consecutive measurements.

**Scenario 4 with TcB measurements (up to 48 h)**
At least three measurements are needed for prediction up to 48 hours. The algorithm requires the initial TcB measurement to be between 8 and 72 hours of postnatal age and at least two additional TcB measurements between 24 and 96 hours of postnatal age, with ≥ 8 hours between two consecutive measurements.

**Scenario 5 with TcB measurements (up to 24 h)**
At least three measurements are needed for prediction up to 24 hours. The algorithm requires the initial TcB measurement to be between 8 and 72 hours of postnatal age and at least two additional TcB measurements between 24 and 96 hours of postnatal age with ≥ 8 hours between two consecutive measurements.

Performance of the NeoPrediX B.1 algorithm was evaluated by applying exactness and clinical acceptance criteria. The exactness criterion was defined as 95% confidence interval (95%-CI) of prediction differences being between +/-70 µmol/l with all differences between +/-85 µmol/l. The clinical acceptance criterion was defined as 95%-CI of differences being between +/-85 µmol/l.

Three different populations were analyzed: A per protocol (PP) population, an intention to treat (ITT) population, as well as a population with extended prediction horizon (modITT). The latter two populations (i.e., ITT and modITT) were chosen as a stress test to challenge the NeoPrediX B.1 algorithm.

**Per Protocol Population**

The PP population was the most rigorous study population defined by the following inclusion criteria: (i) inclusion into the study, (ii) availability of all input parameters required by the NeoPrediX B.1 algorithm, i.e., birth date and time, bilirubin measurement date and time, birth weight, gestational age and delivery mode, (iii) no observed value after start of phototherapy during study, (iv) all bilirubin measurements are compliant with predefined time intervals, i.e., for scenarios 1 and 5 no observed values > 24 hours were accepted, for scenarios 2 to 4 no observed values > 48 hours were accepted.

**Intention to Treat Population**

The ITT population was less rigorous than the PP population as only inclusion criteria (i) to (iii) must be fulfilled.

**Modified Intention to Treat Population**

The modITT population was also less rigorous than the PP population, as it accepted a certain proportion of the newborn infants with an extended prediction horizon. For scenarios 1 and 5 the predefined prediction range is ≤ 24 hours after the 1st measurement, whereas 20% of all infants have an extended accepted prediction horizon of up to 30 hours after the 1st measurement. For scenarios 2 to 4 the predefined prediction range is ≤ 48 hours after the 2nd/3rd measurement, whereas 20% of all infants have an extended accepted prediction horizon of up to 60 hours after the 2nd/3rd measurement.

## RESULTS

In this section, the results of the performance of the NeoPrediX B.1 algorithm are presented for the PP population and the two stress-test populations.

**PP population (n = 179)**

The primary endpoint (i.e., 95%-CI of exactness), as well as the 95%-CI of clinical acceptance criteria were met in all scenarios utilizing TSB. In scenario 4 the 95%-CI was slightly larger than the predefined exactness criteria (75.9 µmol/l compared to 70 µmol/l) while the clinical acceptance criterion were fulfilled. This finding is not surprising as the underlying variability in TcB measurements is larger as in TSB measurements, resulting a slightly increased variability of prediction differences with TcB data. Moreover, there was only a small portion of infants (if at all) with an under-prediction < -85 µmol/l (1.2% in scenario 1, 0.7% in scenario 2 or 3, and 0% in scenarios 4 and 5) or an over-prediction > 85 µmol/l (1.2% in scenario 1, 2.9% in scenario 2 or 3, 5.3% in scenario 4, and 0% in scenario 5).

**ITT (n = 276) and modITT (n = 234) populations**

For the two stress-test populations, predictive capability beyond 24 hours (i.e., up to 30 hours) for scenario 1 and 5, as well as beyond 48 hours (i.e., up to 60 hours) for scenarios 2, 3, and 4 were investigated. The clinical acceptance criterion was fulfilled across all scenarios. For scenarios 1 and 5 also the 95%-CI exactness criteria were fulfilled. In addition, in scenario 4 with TcB measurements the 95%-CI exactness criterion were similar than that of the PP population. In both stress-test populations, number of newborn infants with an under-prediction < -85 µmol/l or those with an over-prediction > 85 µmol/l were low and comparable across all scenarios.

## DISCUSSION

Performance of the NeoPrediX B.1 algorithm was consistent with that in previous validation studies utilizing retrospective TSB data from Regensburg, Germany, and retrospective TcB data from Patras, Greece. Median absolute prediction difference was 8 to 10% with TSB (scenarios 1 and 2), 10 to 14% with TcB (scenarios 4 and 5) across all study populations and scenarios (see Tables 1 and 2). It should be noted that the stress-test populations with extended prediction horizons were included to test the capability of the algorithm beyond the predefined prediction horizons in the NeoPrediX B.1 tool.

As expected, scenarios predicting up to 24 hours (scenarios 1 and 5) performed better than scenarios predicting up to 48 hours (scenarios 2, 3 and 4), and scenarios with TSB measurements performed slightly better than those with TcB measurements. The NeoPrediX B.1 algorithm precisely predicted bilirubin progression in the PP population, with solid performance in the modITT and ITT populations, even though infants with predictions up to 60 hours after the last measurement were included in these stress-test populations.

Overall, study results are consistent with those obtained in previous retrospective studies, confirming that the NeoPrediX B.1 algorithm can accurately predict bilirubin changes in newborn infants, even beyond two days into the future. As such this prospective clinical study demonstrates that the NeoPrediX B.1 tool can support caregivers in personalized and optimized decision making in clinical practice.

| Total serum bilirubin (TSB) measurements | | Prospective study (Tübingen) | | | Retrospective study (Regensburg) | |
|---|---|---|---|---|---|---|
| Scenario | Prediction horizon | Population | Median of absolute prediction difference in % | Median of absolute prediction difference in µmol/l | Median of absolute prediction difference in % | Median of absolute prediction difference in µmol/l |
| Scenario 1 | up to 24 h | PP | **8.7 %** | 16.6 µmol/l | **8.5 %** | 17.4 µmol/l |
| One TSB measurement | up to 30 h | modITT | **8.5 %** | 15.8 µmol/l | **7.9 %** | 15.7 µmol/l |
| Scenario 2 | up to 48 h | PP | **9.2 %** | 17.6 µmol/l | **9.2 %** | 21.5 µmol/l |
| Two serum measurements | up to 60 h | modITT | **9.5 %** | 20 µmol/l | **9.9 %** | 22.3 µmol/l |

**Table 1** Comparison of median absolute prediction difference for total serum bilirubin (TSB) measurements (scenarios 1 and 2) with retrospective study in Regensburg, Germany.

| Total serum bilirubin (TSB) and transcutaneous bilirubin (TcB) measurements | | Prospective study (Tübingen) | | | Retrospective study (Patras) | |
|---|---|---|---|---|---|---|
| Scenario | Prediction horizon | Population | Median of absolute prediction difference in % | Median of absolute prediction difference in µmol/l | Median of absolute prediction difference in % | Median of absolute prediction difference in µmol/l |
| Scenario 3 | up to 48 h | PP | **14.2 %** | 22.4 µmol/l | Not tested in retrospective study | |
| One TSB and one TcB measurement | up to 60 h | modITT | **14.2 %** | 24.2 µmol/l | | |
| Scenario 4 | up to 48 h | PP | **13.8 %** | 26.6 µmol/l | **13.3 %** | 23.1 µmol/l |
| Three TcB measurements | up to 60 h | modITT | **13.8 %** | 26.9 µmol/l | **15.1 %** | 24.8 µmol/l |
| Scenario 5 | up to 24 h | PP | **10.5 %** | 18.5 µmol/l | **10.4 %** | 17.3 µmol/l |
| Three TcB measurements | up to 30 h | modITT | **10.1 %** | 16.7 µmol/l | | |

**Table 2** Comparison of median absolute prediction difference for scenarios with combined TSB and TcB (scenario 3), or TcB-only (scenarios 4 and 5) measurements with retrospective study in Patras, Greece.