Prognostic analysis of road safety in Poland

An update of Appendix I in 'Road safety in Poland', SWOV report R-94-58 (using data from 1953-1993), now based on data from 1953-1995 (incl.) and a partially improved methodology

R-96-41
M.J. Koornstra
Leidschendam, 1996
SWOV Institute for Road Safety Research, The Netherlands

Report documentation

Number:	R-96-41
Title:	Prognostic analysis of road safety in Poland
Subtitle:	An update of Appendix I in 'Road safety in Poland', SWOV report R-94-58 (using data from 1953-1993), now based on data from 1953-1995 (incl.) and a partially improved methodology
Author(s):	M.J. Koornstra
Research manager:	F.C.M. Wegman
Project number SWOV:	69.048
Keywords:	Safety, improvement, data bank, forecast, danger, traffic, fatality, injury, mathematical model, policy, evaluation (assessment), Poland.

Contents of the project:	In 1994, the development of traffic as measured by passenger cars and road safety as measured by road fatalities has been analysed on the basis of the data from 1953 to 1993 inclusive for Poland by SWOV, using models for traffic growth and risk development. That analyses tentatively resulted in alternative prognoses for passenger cars, fatality risk per 1,000 cars and fatalities. The data for 1994 and 1995 are now available; on the basis of these, the prognostic analysis is updated.

Number of pages:	12 pp. + 3 pp.
Price:	Dfl. 15,-
Published by:	SWOV, Leidschendam, 1996

SWOV Institute for Road Safety Research
P.O. Box 170
2260 AD Leidschendam
The Netherlands
Telephone 31703209323
Telefax 31703201261

Contents

1. Introduction

The development of traffic as measured by passenger cars and road safety as measured by road fatalities has been analyzed on the basis of the data from 1953 to 1993 inclusive for Poland by SWOV using their models for traffic growth and risk development (Koornstra, Mulder & Wegman, 1994). That analyses tentatively resulted in alternative prognoses for passenger cars, fatality risk per 1,000 cars and fatalities.

The data for 1994 and 1995 are now available. A prediction and outcome comparison shows that the model for passenger car growth, both the high and low growth alternatives, predicted in the short term a higher number of cars than observed for 1994 and 1995 (low prediction: 7,295 and 7,934 thousand; high prediction 7,792 and 8,767 thousand; observed are 7,153 and 7,500 thousand cars).

The observed fatality rates per thousand cars for 1994 and 1995 are 0.943 and 0.953, where as the risk model alternatives predicted rates of 0.868 and 0.763 for the prognostic low risk alternative and 0.924 and 0.845 for the prognostic high risk alternative. So both risk model alternatives underestimated the actual short term risk development.

The combined alternatives of high growth prognosis and high risk prognosis predicted fatalities for 1994 and 1995 of 7196 and 7,409, where the actual fatalities are 6,744 and 7,150. The increase is well predicted but the predicted level for the short term prognosis obviously is too high. The model combinations for low risk prognosis predicted a decrease in fatalities from 1994 to 1995 and underestimated the actual fatalities. Due to the compensatory effects of under- and overestimation in the growth and risk models the mean of the combinations for high growth times low risk and low growth times high risk would have given a fairly good prediction.

In the time passed SWOV has improved the growth model. Apart from cyclic modulations for deviations, the model was based on symmetric logistic growth. The improved model allows for asymmetric logistic growth (asymmetric S-curves, including the Gompertz curve as the limit case for the most left asymmetric S-curve with fast starting growth and a slow levelling-of). Since the asymmetry and the periods for cyclic deviations are interdependent, the mentioned short term prediction errors for growth of cars might be caused by a too long upward cyclic deviation pattern around the rigid symmetric S-curve for growth fitted to the earlier data up to 1993. This may explain the overestimation in both the low and high prognostic alternatives for growth of passenger cars.

The monotonic steep risk decay observed in Poland up to 1988 is suddenly disturbed by marked higher risks between 1989 and 1992. Since thereafter a lower risk than 1988 only was known for 1993, the predictive risk analysis from 1994 probably could estimate the recent pattern for risk deviations around exponential risk decay sufficiently reliable. Now there are two recent years with risk values to add (with rather small risk reductions), it may enable a more precise estimation of the influence of the recent disturbance of the monotonic risk decay on the deviation pattern of the future risk deviations from the underlying exponential risk decay.

2. Updated and revised prognoses

2.1. Traffic growth analysis

In the analysis of 1994 (for data up to 1993) we a priori have determined alternative saturation level of passenger car growth as 1 car per 2 or 3 inhabitants. By the original model it is not possible to estimate the saturation level for growth from the data if the given time-series of growth data has not reached more than 50% of the saturation level. In principle this remains a problem, but the improved growth model with its asymmetric growth (closer fit to observed data and possible earlier inflexion points of underlying asymmetric S-curves) may be able to estimate more sensibly a saturation level from shorter series of data (for example the saturation level of the Gompertz curve already can be determined after more than 37% of the maximum growth is reached). Since for different saturation levels the deviations from fitted growth are different, the estimated saturation level very much depend on the minimization of the error distribution for the growth data. Different error minimizations may give different maxima for the same time-series, if no data for many years are available after the time of the implied inflexion point of the growth curve.

It happens that the least square minimization of raw deviations determines an underlying asymmetric growth for Poland that is a Gompertz curve (as it does for many other countries) with a maximum of 22,367 thousand passenger cars and an inflexion point of the underlying S-curve around 1998, modulated by harmonic cyclic deviations with periods of 30, 15 and 7.5 years. This maximum is not very much different from the higher prior assumed maximum of 20 million cars from the prognostic analysis in 1994 (1 car per 2 inhabitants in the future). If we minimize the Chi^2 of the deviations the fitted asymmetry determines an underlying S-shape that is nearly half way between a Gompertz and a symmetric S-curve (asymmetry parameter is 0.549) with a maximum of 12,888 thousand cars and its inflexion point around the year 1991, while the harmonic deviation cycles have periods of 32, 16 and 8 years. This latter maximum happens to be close to the lower prior assumed maximum of 13 million in the prognosis made in 1994 (1 car per 3 inhabitants in the future). Therefore, we display in *Figure 1* these two updated and revised prognoses as replacements for the alternatives from the prognosis made in 1994 for traffic growth.

If we minimize the logarithmic or proportional deviations, which seems most appropriate if growth data exhibits a constant coefficient of variation, as traffic volumes and data with ratio-scales (positive values with a meaningful zero scale point) generally show, the estimated maximum becomes about 11 million cars. This latter level of maximum cars seems rather low in view of the recent Polish traffic growth and the generally expected levels of motorization in industrialized countries. Therefore, the model outcomes for these error minimizations are disregarded.

The prognoses from a least square and a minimum Chi^2 solution for growth of cars even may be defended if the true error distribution is indeed characterized by a constant coefficient of variation. In that case the least square solution implicitly weights the most recent years more (and beginning years of the time-series less), which may give more confidence for extrapolation in the future. The minimum Chi² solution for traffic growth may be quite appropriate on its own, since some algebra shows that the in this way estimated growth multiplied by the observed fatality rate minimises the Chi² of the resulting estimates of fatalities. It fits so to speak the traffic growth in order to optimise the prediction of the fatalities.

It is evident from *Figure 1* that for both alternatives the retrospective prediction of the growth of passenger cars in Poland is excellent. The proportion of explained variance for the high alternative is 0.999403 and for the low alternative 0.999365. One may prefer the higher alternative, because its underlying Gompertz S-curve contains one parameter less than the underlying asymmetric logistic S-curve for the lower alternative.

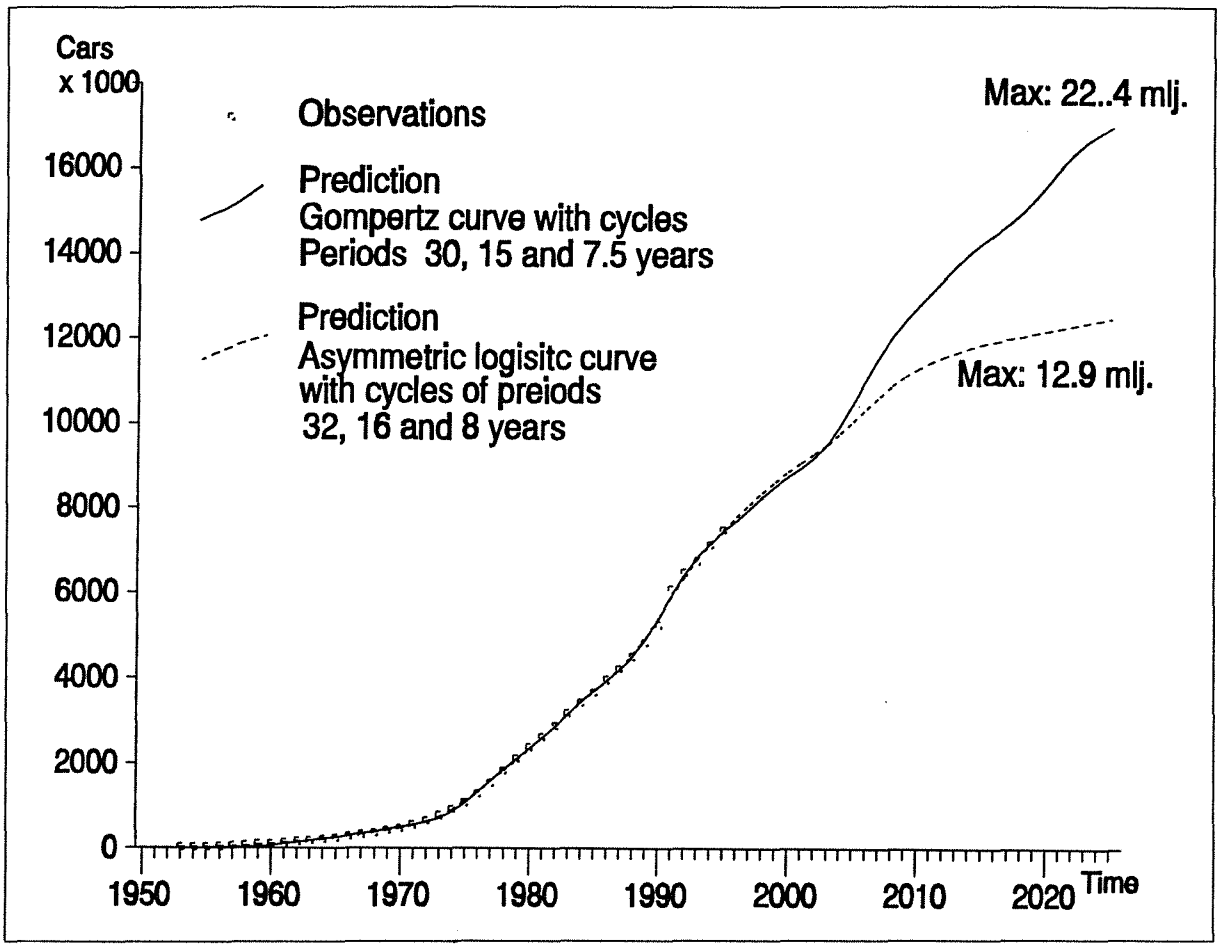

Figure 1. *Updated and revised analysis of passenger cars.*

The alternative prognoses up to about 2003 do not differ very much. Compared to the analysis made in 1994 we now find, due to the new data added and the asymmetry of the underlying S-curves, shorter harmonic cycles around the asymmetric S-shaped main trends. Instead of cycles with periods of nearly 40, 20 and 10 years the best fitting harmonic cycles now have periods of 30, 15 and 7.5 years around Gompertz growth (most asymmetric S-curve for the high prognostic alternative) and periods of 32, 16 and 8 years around the less asymmetric logistic curve (the lower prognostic alternative). Clearly the more left asymmetric the fitted underlying S-curve is the shorter the periods of the deviation cycles become. A somewhat larger influence comes from the 15 or 16 years cycles than from the longest cycles of 30 or 32 years, while the contribution of the shortest cycles of 7.5 or 8 years is only minor. The underlying asymmetric S-curve of the then still steep increasing Gompertz curve with the higher growth maximum is compensated by 13% decrease of the proportional cyclic influences in the period between 1995 and 2003 (of course after 2003 again the cyclic influences become increasing above the still increasing Gompertz curve), while in the same period the then already less steep increasing underlying S-curve of the asymmetric logistic curve for the lower growth alternative is only compensated by 6% decrease of comparable cyclic influences in that alternative with the lower maximum growth. In 2025 the lower alternative has nearly reached its estimated maximum level of passenger cars, whereas the higher alterative prognosis then still envisages a further growth of about 6 million passenger cars.

2.2. Risk analysis

The updated risk decay analysis for the fatality rate per 1,000 passenger cars gives very similar results compared to the analysis made in 1994. The main difference is a now added fourth cycle that becomes with the added data of 1994 and 1995 important enough to include. As in the analysis made in 1994 two alternative models are solved which do not differ very much retrospectively, but do differ markedly in their prognoses. In *Figure 2* the results are shown, where the inlay enlarges the graphic presentation of the analysis after 1980.

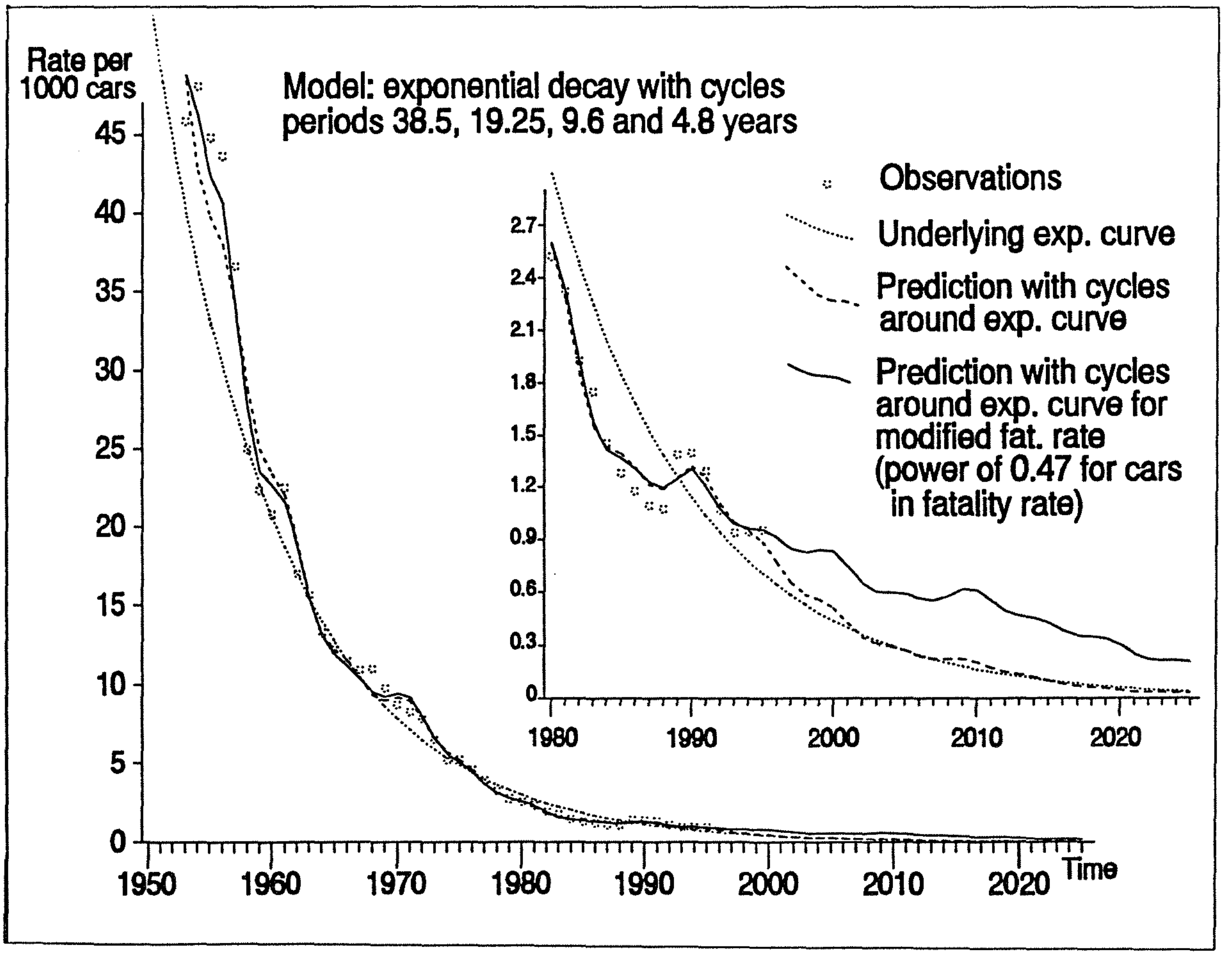

Figure 2. *Updated risk analysis for Poland.*

The periods of the main harmonic cycles for risk deviations around exponential decay are for both alternatives now marginally shorter, that is 38.5, 19,25, 9.6 and 4.8 years, where the contribution of the two shorter cycles is about half of the longer two cycles. The combination of the two shorter cycles contributes mainly in a better fit of risks in the period around 1989-1990 (and of course every 9.6 years before, notably also around 1970-1971 and 1962). The higher prognostic alternative fits an exponential decay (and deviation cycles around it) for a modified fatality rate, that is for fatalities per power-transformed thousands of annual passengers cars instead of per thousand passenger cars it self (or equivalently: fatalities raised by a power factor per thousand cars). Therefore, this alternative needs one parameter more than the more simple alternative with the lower risk prognosis. The estimated power factor on cars is now 0.46846, that is only marginally higher than 0.44878 in the 1994-analysis.
As *Figure 2* shows both alternatives follow in the retrospective prediction the actual risk developments rather closely. Both alternative analyses do describe some risk increase from 1988 to 1990 as well as the risk reduction from 1991 to 1993, but both also imperfectly describe the actual larger risk reductions from 1985 to 1987.

The powered solution for the modified fatality rate (the alternative with high risk prognosis) has now Chi^2=806.5 and for the simple solution (the alternative with low risk prognosis) the Chi^2=960.2. The improved Chi^2 of the new solutions are mainly due to the added fourth cycle (in the analysis made in 1994 with two observations for 1994 and 1995 less the Chi^2 of the corresponding solutions were 988 and 1062). The difference in Chi^2 between the two alternatives has increased compared to the analysis made in 1994, but still does not reach a 0.20 significance level (F-test=1.19 with 31 and 32 df's). Therefore, the question which alternative of the prognoses may be more valid remains unanswered, but in contrast to the two traffic growth alternatives the two risk prognoses remain to differ already in the near future. However, not only the fit of the alternative with higher prognosis is somewhat better, also at face value this alternative may be more likely since the extrapolated risk prediction for the far future becomes unbelievable low for the lower risk alternative (less than 100 fatalities in 2050, where as the alternative with the higher prognosis predicts 2,414 or 3,004 fatalities in 2050, depending on low or high traffic growth). Therefore, on the one hand one may have a slight preference for the higher risk prognosis. Nevertheless on the other hand, as the pictured exponential risk curve (without cycles and power) shows, the risk reduction is in a macroscopic sense rather well characterized by such a steep risk reduction for the past in Poland and so it could be real for its future.

2.3. Analysis of fatalities

The importance of a comparable steep risk reduction as in the past for Poland, again as in the analysis made in 1994, becomes clear from the prognoses of the resulting fatalities. Multiplying the two alternatives for traffic growth with the two alternatives for the risk development by definition yields four alternative descriptions and prognoses for the development of fatalities in Poland. Again, as in the analyses made in 1994, the retrospective fits for the alternatives are not significantly different, as already could be expected from the hardly different fits in *Figure 1* for traffic growth and *Figure 2* for risk developments. In *Figure 3* we picture the results for all the four alternative predictions of fatalities and their prognosis up to 2025. As might be expected from the excellent fit for traffic growth and fairly good fit of the risk developments, the fatalities of the past are well predicted, including the peak in 1991. The actual higher number of fatalities in 1968 and the actual lower fatalities between 1985 and 1989 are the main observed departures from the predicted fatalities. *Figure 3* clearly shows that the prognostic difference in the alternatives for traffic growth are less important than the prognostic differences for the risk alternatives. Compared with the analysis made in 1994 the predicted peaks in fatalities from the higher risk alternative are less high.

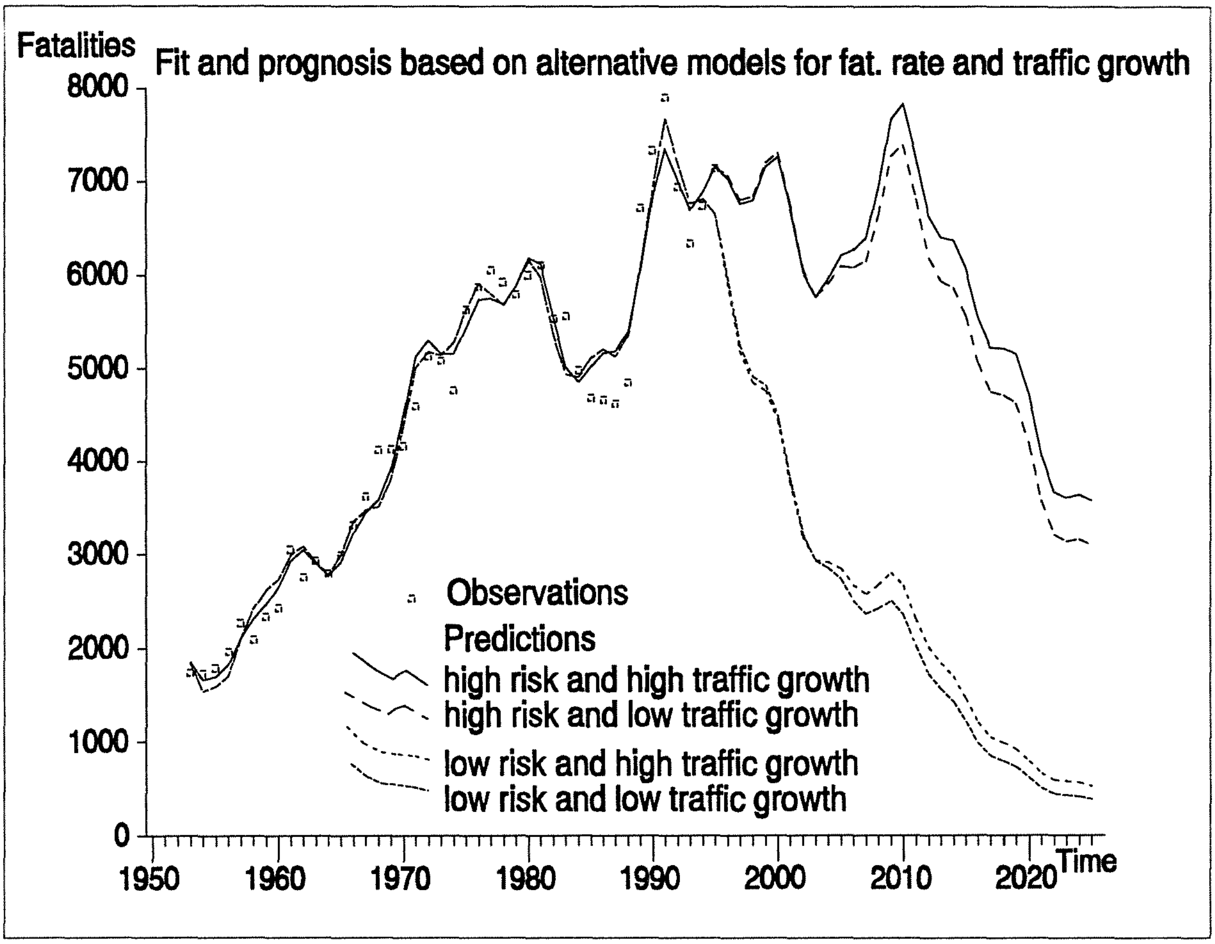

Figure 3. *Analysis of fatalities in Poland.*

The main difference with the analysis made in 1994 for comparable alternatives is that no longer fatalities as high as 9,500 are predicted. The alternative of low risk prognosis combined with each alternative for traffic growth predicts a rather steep and lasting decrease in fatalities, even steeper than in the analysis made in 1994. This decrease is only stagnated around 2008 to 2010 at a temporary level between 2,400 and 2,800 fatalities. By the low risk prognosis the fatalities are predicted to decrease further to even lower than 1,000 after 2015 or 2018, depending on the combination with the growth alternative. The high risk alternative prognosis combined with each traffic growth alternative yields fluctuating fatalities around 7,000 up to the year 2000 and then is followed by a decrease to below 6,000 fatalities around 2003 to 2006 as well as by a marked rise of fatalities towards a peak of about 7,400 to 7,850 (depending on the growth prognosis) in 2010. Thereafter also a lasting decrease in fatalities is observed from growth combined with the high risk prognosis.

In terms of the SWOV-report from 1994 the turning point in time for the road safety in Poland now becomes solely dependent on the risk development. For the lower risk prognosis this turning point is already passed now, whereas for the higher risk prognosis this turning point in time is still located around 2010. The absolute difference in the main prognostic developments is maximal around 2010 with a difference of about 5,000 fatalities per year. It must be concluded that the future of road safety in Poland depends solely on the achievable risk reduction. An effective road safety policy, therefore, is crucial and of national importance.

Literature

Koornstra, M. J., Mulder, J.A.G. & Wegman, F.C.M. (1994). *Road safety in Poland*. SWOV report R-94-58. SWOV, Leidschendam.

Mathematical appendix

The general asymmetric logistic growth model is written as:

$$\mathbf{V_t = V_{max}\{1 + \exp[a.t + b + \sum_{i=1}^{i=n} f_i.\cos(\sigma_i(t - \tau_i))]\}^{-1/c} + e_t} \qquad \mathbf{(1a)}$$

where if $\lim c \to 0$ the expression reduces to the Gompertz growth model as

$$\mathbf{V_t = V_{max}.\exp\{-\exp[a.t + b + \sum_{i=1}^{i=n} f_i.\cos(\sigma_i(t - \tau_i))]\} + e_t} \qquad \mathbf{(1b)}$$

and where

V_t	= observed traffic volume or amount of passenger cars in year t
V_{max}	= saturation level of volumes or cars
a	= slope or time scale parameter of S-curve
b	= time location parameter of S-curve
c	= asymmetry parameter of S-curve
f_i	= weight or amplitude parameter for cosine cycle i
τ_i	= location or phase parameter for cosine cycle i
σ_i	= frequency or period length parameter for cycle **i**, where the the parameters are constraint to harmonic cycle frequencies with $\sigma_i = 2.\sigma_{i-1}$
n	= number of harmonic cosine cycles
e_t	= V_t - V_t = error term for observed value in year t
V_t	= estimated traffic volume or cars in year t

The general fatality or risk model is written as

$$\mathbf{F_t = V_t^{-s}.\exp[\alpha.t + \beta + \sum_{j=1}^{j=m} g_j.\cos(\pi_j(t - \mu_j))\} + \epsilon_t} \qquad \mathbf{(2)}$$

where for s=1 the fatality rate F_t/V_t is fitted as an exponential decay function of time with proportional cyclic deviations and where

F_t	= observed road fatalities in year t
α	= slope or time scale parameter of exponential decay curve
ß	= time location parameter of exponential decay curve
V_t	= observed traffic volume or amount of passenger cars in year t
s	= power parameter of V_t;
g_j	= weight or amplitude parameter for cosine cycle j
μ_j	= location or phase parameter for cosine cycle j
π_j	= frequency or period length parameter for cycle j, where the the parameters are constraint to harmonic cycle frequencies with $\pi_j = 2.\pi_{j-1}$
m	= number of harmonic cosine cycles
ϵ_t	= F_t - F_t = error term for observed value in year t
F_t	= estimated fatalities in year t

For the least square minimization of e_t the growth data in thousands of cars from 1953 to 1995 in Poland the minimized sum of squared errors SSe becomes SSe = 105843.355 for the Gompertz curve with c -> 0 and with t = 1953, 1954 ... 1995 inclusive the parameter values are:

V_{max} = 22367.5
a = -0.04652
b = 92.9484
f_1 = 0.03293
f_2 = 0.03706
f_3 = 0.01035
τ_1 = -1.26836
τ_2 = -0.10720
τ_3 = 0.77916
σ_1 = 0.20944 (cycles of 30, 15 and 17.5 years)
n = 3

For the Chi-square minimization of e_t the growth data in thousands of cars from 1953 to 1995 in Poland the minimized Chi²= $\Sigma(V_t-\mathit{V}_t)^2/\mathit{V}_t$ = 31.81 and with t = 1953, 1954 ... 1995 inclusive the parameter values are:

V_{max} = 12887.6
a = -0.10203
b = 202.5324
f_1 = 0.04255
f_2 = 0.06786
f_3 = 0.01329
τ_1 = 0.24337
τ_2 = 1.35071
τ_3 = 0.22740
σ_1 = 0.19635 (cycles of 32, 16 and 8 years)
n = 3

For the Chi² minimization of ϵ_t with s fixed to unity, that is the exponential decay function with cycles fitted to the fatality rate the minimized Chi²= $\Sigma(F_t-\mathit{F}_t)^2/\mathit{F}t$ = 960.21 for Poland and with t = 1953, 1954 ...1995 inclusive the parameter values are:

α = -0.096418
ß = 192.005129
s = 1.0 (a priori fixed)
g_1 = 0.160308
g_2 = 0.166948
g_3 = 0.047012
g_4 = 0.053201
μ_1 = -0.921852
μ_2 = 0.724530
μ_3 = -1.880500
μ_4 = 0.270497
π_1 = 0.163200 (cycles of 38.5, 19.25, 9.625 and 4.8125 years)
m = 4

For the Chi^2 minimization of ϵ_t with s also estimated, that is the exponential decay function with cycles fitted to the modified fatality rate the minimized Chi^2= $\Sigma(F_t-\mathit{F}_t)^2/\mathit{F}t$ = 806.54 for Poland and with t = 1953, 1954 ...1995 inclusive the parameter values are:

α = -0.026910
β = 58.347599
s = 0.468459
g_1 = 0.126909
g_2 = 0.113688
g_3 = 0.065009
g_4 = 0.045927
μ_1 = 5.054515
μ_2 = 1.394827
μ_3 = -2.218013
μ_4 = 0.458351
π_1 = 0.163200 (cycles of 38.5, 19.25, 9.625 and 4.8125 years)
m = 4

The prognosis for the fatalities and the fatality rates are obtained by replacing in (2) V_t by V_t obtained from (1a) or (1b) for years after 1995.