Supplementary material to article by M. Gillstedt and S. Polesie "Ability to Predict Melanoma Within 5 Years Using Registry Data and a Convolutional Neural Network: A Proof of Concept Study"

Appendix S1

Description of the original research cohorts

The original study cohorts consisted of eight cohorts (four disease and four control cohorts) used in previous research projects (registration number 911-17 approved by the regional ethical review board in Gothenburg).

- The first cohort consisted of patients with any dispensed methotrexate (Anatomical Therapeutic Chemical [ATC] codes L04AX03 and/or L01BA01) from Swedish pharmacies in the period 1 July 2005 to 31 December 2016. This cohort included 124,520 individuals.
- 2. The second cohort consisted of patients with any dispensed prescription of tumor necrosis factor alpha inhibitor (ATC codes: L04AB01, L04AB02, L04AB04 or L04AB05, L04AB06) and/ or ustekinumab (ATC code L04AC05) from Swedish pharmacies in the period 1 July 2005 to 31 December 2016. This cohort included 42,536 individuals.
- 3. The third cohort consisted of patients who had ≥2 outpatient diagnoses of psoriasis (International Classification of Diseases 10th Revision code L40.0-L40.9) in the period 1 January 2001 to 31 December 2016. At least one diagnosis needed to be made by a dermatologist. This cohort included 81,738 individuals.
- 4. The fourth cohort consisted of patients with any dispensed prescription with calcipotriol and/or combination of calcipotriol (ATC codes D05AX52 and/or D05AX02) from Swedish pharmacies in the period 1 July 2005 to 31 December 2016. At least 1 diagnosis needed to be made by a dermatologist. This cohort included 81,619 individuals

Cohort used in the present investigation

To each of the cohorts above, four additional cohorts each consisting of five age-, sex and geographically matched individuals per case, drawn randomly from the Swedish population, were available. To exclude any systematic bias (i.e. to avoid an overrepresentation of patients with psoriasis) we only included the control cohorts in our analyses. When merging all control cohorts 1,489,519 individuals were available. The study flow-chart is demonstrated in Fig 1.

Appendix S2

Background to Swedish healthcare and used databases/registries

Swedish citizens have universal access to healthcare. At birth or when gaining Swedish citizenship all inhabitants receive a unique personal identification number that facilitates linking between databases and registries. The Swedish registers and databases related to healthcare are administered by the National Board of Health and Welfare and the population registers are administered by Statistics Sweden. In the following sections a brief description of the healthcare registries and databases is presented. All variables used in our investigation (time-dependent and time independent) are described in Tables I and II.

The Prescribed Drug Register contains information on all filled prescriptions from Swedish pharmacies from the start of the register in July 1, 2005. The estimated coverage is close to 100%. Drugs dispensed in an inpatient setting is not covered by the register. For this investigation the date of the redeemed drugs and the Anatomical Therapeutic Chemical (ATC) classification code were used. The ATC is an internationally accepted classification system for medicines that is maintained by the WHO. Each drug is given a specific code with seven symbols. The first symbol is a letter that represents the main group. The second and third numbers both denote the therapeutic group. The fourth number denotes the pharmacological group. The fifth symbol denotes the chemical subgroup and the sixth and seventh numbers represents the active substance.

The Cancer Register includes all incident cancers (including in situ cancers) from 1958. In Sweden cancer reporting is mandatory both for the responsible physician and reporting pathologist, the estimated coverage is high (i.e. >95%). For this investigation the dates of cutaneous malignant melanomas (including in situ melanoma) were obtained.

The Cause of Death Register includes dates and causes of death for all diseases citizens. Only individuals that were alive at December 31, 2011 were included. Individuals that died in the period 1 January 2012 to 31 December 2016 were included.

The Outpatient Register includes information on diagnoses in non-primary outpatient care, coded according to the International Classification of Diseases, 10th revision (ICD-10) from 2001. For this investigation all ICD-10-codes and the dates for each diagnosis was used.

The Inpatient Register covers inpatient hospitalizations (in Sweden) for all Swedish citizens from 1964. The date of admission and the date of discharge as well as the main diagnoses and all other diagnoses are recorded. For this investigation all ICD-10-codes and the admission date for each diagnosis were used.

The Population Register includes data on residency and dates of immigration and emigration for all people residing in Sweden, and coverage is virtually complete. Data on educational level, income level, civil status, geographical region and country of birth are included if those data are available. For this investigation we used one imputation (last known) for most variables. Only individuals that had no migration events in the period 4 July 2005 to 31 December 2016 were included.