Appendix. Data source of the population: the SNIIR-AM

Data source is the *Système National d'Information Inter-Régimes de l'Assurance Maladie* (SNIIR-AM) that is the unique database of the French National Health Insurance System. It is handled by the *Caisse Nationale de l’Assurance Maladie des Travailleurs Salariés* (CNAMTS) that is the organization in charge of health care reimbursements. It collects demographic and health expenditure reimbursements data of the entire French population (65,586 million inhabitants in January 2013). These data are individualized, anonymous, exhaustive, and linkable for a given patient. Therefore, the SNIIR-AM has been used since the mid-2000s to conduct large pharmacoepidemiological post-authorization studies. Briefly, it is constituted by several datamarts colligating prospectively for each patient the following data:

- In the administrative datamart: age, gender, department and town of residence, vital status (date of death if applicable), and insurance scheme.

- In the costly long-term disease (*Affections de Longue Durée*, abbreviated ALD) datamart: ALD list. ALD allows full reimbursement of every health care related to that disease. They are encoded with the International Classification of Diseases, version 10 (ICD-10, available at: http://www.who.int/classifications/icd/en). Dates of start and dates of end are recorded. Occupational diseases are also recorded in this datamart.

- In the out-hospital drug reimbursement datamart: this datamart contains the date of dispensing, the drug name and the dosage form encoded with the *Club InterPharmaceutique* classification (available at: http://www.ucdcip.org) and the quantity (number of boxes) dispensed. Prescribed dose and indication are not available. Over-the-counter dispensed drugs are not recorded because there are not reimbursed.

- In the out-hospital biology datamart: it contains the date of sampling, names of dosage performed encoded with the Nomenclature des Actes de Biologie Médicale (available at: http://www.ameli.fr/accueil-de-la-ccam/index.php). Results of lab tests are not recorded.

- In the hospitalization datamart, called the Programme de Médicalisation des Systèmes d’Informations (PMSI): this datamart contains data of all hospital stays in public and private hospitals. Data are: entry and discharge dates, age at entry date, identifier hospital code, whether the patient was admitted to intensive care unit, diagnoses (for each stay: one primary diagnosis, one related diagnosis and all the associated diagnoses) encoded with the ICD-10, expensive drugs (names and dosage) dispensed during the stay encoded with the Unités Commune de Dispensation (UCD) classification (available at: http://www.ucdcip.org) with the quantity (number of boxes) dispensed. All medical and surgical procedures and interventions are also encoded using the Classification Commune des Actes Médicaux.

- Sick-leave datamart with dates of start and of end.

- All these information are linked thanks to the patient identification number (Numéro d’Inscription au Répertoire, abbreviated NIR) which is the unique number identifying a given adult patient. Children of a given adult patient have the same NIR than the adult parent who gives them the right to benefit from the national Health Insurance System.

According to French law, available data in the current database are those of the current year and of the three previous years (Arrêté du 11 juillet 2012 relatif à la mise en œuvre du système national d’information inter-régimes de l’assurance maladie. Journal Officiel de la République Française 2012. Available at: http://www.legifrance.gouv.fr/
Data extraction for the French Adult primary Immune Thrombocytopenia: a pHarmacoepidemiological study (FAITH)

In end 2012, the CNAMTS computer engineers extracted the data for all patients encoded for an ALD or a hospital stay with the ICD-10 code related to ITP (D69.3) from the 1st January 2009 to the 31st December 2011. These data were delivered in 2013 on a CD-ROM support to our research unit. Annual extractions until 2022 are foreseen using the same process, along with extraction of data regarding the patients included in the cohort the previous years.

All the data recorded in the SNIIR-AM datamarts for these patients are delivered, except patients’ NIR. As a result, delivered data are made fully anonym. Patient linkage through datamarts is performed thanks to a study number attributed by the CNAMTS computer engineers, which has been elaborated for a given patient based on its NIR, its year of birth and its gender. As a result, these study numbers differentiate children from their parents. Twin children are differentiated thanks to another variable encoding the twinning rank.
Figure S1. Monthly incidence of ITP by age groups.