Health Services Research: access to, and integration of, inpatient and outpatient health administrative data for assessing healthcare systems' performance

Case study 2: Evaluation of the impact of health policy decisions on the utilization of treatment services by individuals with and without chronic diseases (e.g., cancer) for comparison.

on behalf of the Federal Office of Public Health (FOPH)

contract ID:142006407

Marie-Annick Le Pogam[1], Murielle Bochud[1]

[1]Department of Epidemiology and Health Systems, Unisanté, University Center for Primary Care and Public Health, Lausanne, Switzerland

DOI : https://doi.org/10.16908/pub.2025.004

This report was produced by Unisanté in the context of a mandate (Proof-of-concept-DigiSanté) including 3 case studies as part of the DigiSanté program. This case study illustrates the secondary use of health-related data, as supported in the DigiSanté project "Datenraum für die gesundheitsbezogene Forschung" (DigiSanté Programm zur Förderung der digitalen Transformation im Gesundheitswesen, package 4). It relates to data use categories (1) Academic research and (4) Scientifically elaborated information for stakeholders in healthcare system and in health policy (e.g., authorities).

Context

The research project presented in this report was funded by the Swiss Cancer Research Foundation (grant n° HSR-5225-11-2020) and received ethical approval from the Cantonal Commission for the Ethics of Research on Human Beings (CER-VD) in Lausanne, Switzerland.

The COVID-19 pandemic significantly disrupted cancer care services worldwide, primarily due to the reallocation of healthcare resources. These disruptions led to widespread delays, postponements, and

cancellations of services, including ambulatory care, hospital admissions for elective treatments, and chemotherapy transport. At the same time, the crisis exposed opportunities to address inefficiencies such as over-diagnosis, over-treatment, and resource overuse, and spurred innovations like telemedicine, national task forces, and new models of cancer care delivery.

In Switzerland, however, research on the impact of the pandemic on cancer care, patient outcomes, healthcare service utilization, and related costs remains limited. To fill this gap, we conducted the project titled *"Impact of the COVID-19 crisis on the quality of cancer care in Switzerland: a controlled time-series analysis using insurance and hospital claims data (TOCCATA)"*. This study investigated:

1. The impact of public health policy measures - specifically the lockdown during the COVID-19 crisis - on inpatient and outpatient care utilization among cancer and non-cancer patients.
2. The extent to which the lockdown affected the continuity of care for these patients.
3. The health and financial consequences of the crisis, focusing on mortality, hospital readmissions, length of stay, healthcare-related adverse events, and healthcare costs.
4. Whether cancer patients' experiences, as captured through the SCAPE2 study[a] questionnaires, aligned with observed patterns in inpatient and outpatient healthcare data. This objective sought to assess the consistency between patient-reported experiences in cancer care and health outcomes derived from administrative health data.

As outlined in the protocol, we analyzed administrative health data from adult cancer and non-cancer patients in Switzerland for the period 2017–2021. Data sources included insurance claims data, as well as inpatient and outpatient administrative health data from hospitals. Additionally, we used aggregated data from the SCAPE2 study to explore correlations between healthcare utilization or health outcomes and reported experiences of cancer patients during the COVID-19 pandemic. This research aimed at guiding policymakers and healthcare stakeholders in developing strategies to mitigate the impact of future health crises on cancer care.(1)

Despite its contributions, the study faced significant challenges related to data access, processing, and linkage, compounded by ethical, regulatory, and administrative constraints. These issues caused delays, limited the scope of the analyses, increased costs and workloads, and ultimately impacted the dissemination and relevance of the findings.

Recognizing that similar challenges may affect many health services research projects relying on administrative health data, this case study seeks to share lessons learned and provide recommendations to enhance the usability and integration of such data within a national data ecosystem. These insights are intended to align with the goals of the DigiSanté program and to support the digital transformation of Switzerland's healthcare system.

Data sources and characteristics

Administrative health data, also referred to as **billing data**, is a broad term encompassing data routinely collected primarily for the purposes of healthcare administration, reimbursement, and financial accounting within health systems.

These data include patients' demographics, diagnostic, procedural, and medication codes, service details, as well as charges and payment information, among others.

Specifically, the administrative health data that were analyzed within the research project consisted of:

[a] An overview of the SCAPE2 study can be found here: Patient experiences with cancer care.

1. Swiss hospital medical statistics data (*StatMed or MS* data) for inpatient care provided by the Federal Statistical Office (FSO).
2. Swiss hospital outpatient data (*PSA* data) for hospital outpatient care provided by FSO.
3. Swiss economic data statistics by case (*FKS* data) provided by FSO.
4. Health insurance claims data on hospital-based and community-based outpatient care, provided by two major Swiss health insurers.

1- *StatMed dataset* *(Statistique médicale des hôpitaux, Medizinische Statistik der Krankenhäuser, Statistica medica ospedaliera)*

This *StatMed* [b] dataset provides comprehensive national data on inpatient (stationary) hospital stays in Switzerland. Managed by the FSO, these medical statistics have been collected since 1998 in accordance with the Federal Statistics Act (*SR 431.01 BStatG/FSA*) and the Federal Health Insurance Act (*SR 832.10 KVG/LAMal*). Since 2012, data collection has been mandatory on an annual basis for all Swiss hospitals, clinics, maternity units and birth houses.

This dataset includes detailed information on providers, patients, and stays[c], including hospital type or canton, patient sex and age at admission, length of stay, month of admission, admission and discharge mode, and in-hospital mortality. Clinical diagnoses are coded using ICD-10-GM[d], while procedures are recoded using the CHOP[e] classification. *StatMed* data are extracted from discharge letters and all available information in patients' electronic medical records (EMR)[f] in a standardized, structured, and anonymized manner. As such, it serves as a key resource for analyzing hospital utilization, inpatient outcomes, healthcare delivery, and trends within the Swiss healthcare system.

In the TOCCATA study, *StatMed* was used to analyze patterns of hospitalization, inpatient outcomes, and the effects of the COVID-19 lockdown on cancer and non-cancer patient populations. The dataset enabled longitudinal analyses of key indicators such as length of stay, mortality, and hospital readmissions.

2- *PSA dataset* (*Patientendaten Spital Ambulant, Données des patients ambulatoires des hôpitaux, Dati dei pazienti del settore ambulatoriale ospedaliero*)

The *PSA* dataset specifically refers to hospital outpatient data, covering all ambulatory services provided within Swiss hospitals and maternity units (2). *PSA* data collection has been mandatory for all hospitals in Switzerland since 2017, in accordance with the *SR 431.01 BStatG/FSA* and the *SR 832.10 KVG/LAMal*.

This dataset covers all hospital-based outpatient services that do not require an overnight stay, spanning general medicine, specialty services, minor surgeries, diagnostic imaging, and therapy sessions (2). PSA data are derived from billing records coded according to various tariff structures, depending on the type of service provided. The most commonly used billing systems include (2):

- TARMED – The standard tariff for medical and technical outpatient services, covering physician consultations, diagnostic tests, minor procedures, and imaging.

[b] Fiche signalétique : Statistique médicale des hôpitaux

[c] Hospital medical statistics - Medical statistics variables : Statistique médicale des hôpitaux | Office fédéral de la statistique (admin.ch)

[d] ICD-10-GM: International Statistical Classification of Diseases and Related Health Problems 10th Revision German modification (Instruments pour le codage médical / Instrumente zur medizinischen Kodierung / Strumenti di codifica medica)

[e] CHOP: Swiss Classification of Surgical Procedures (Instruments pour le codage médical / Instrumente zur medizinischen Kodierung / Strumenti di codifica medica)

[f] See eHealth glossary for the definition of electronic medical record (EMR): Glossaire - eHealth Suisse : Définitions & explications / Glossar - eHealth Suisse: Definitionen & Erklaerungen / Glossario – eHealth Suisse: definizioni e spiegazioni

- Pharma Code – Used for billing outpatient medications and pharmaceuticals, ensuring the correct reimbursement of prescribed drugs.
- Hospital Tariff for Other Ambulatory Services – Applied to non-TARMED hospital outpatient services, including certain nursing, therapeutic, or technical procedures not covered by TARMED.
- Federal List of Analyses (*Liste des analyses*, *LA*) – Specifies the reimbursable laboratory tests, including blood tests, microbiological analyses, and genetic diagnostics.

Other specialized tariffs, such as those for physiotherapy, dialysis, home care services, and accident insurance (*LAA/UVG*), may also apply depending on the type of outpatient care provided.

Each record in the PSA dataset corresponds to a unique outpatient encounter, enabling data linkage and efficient processing of large volumes of information for statistical and longitudinal analyses. Due to its national scope and high level of standardization, PSA serves as a vital resource for monitoring hospital-based ambulatory care, analyzing healthcare expenditures, and supporting evidence-based health policy in Switzerland. It plays a central role in health services research and system monitoring, particularly in evaluating service provision, cost structures, and utilization trends within hospital outpatient settings. However, the PSA dataset does not capture care delivered outside of hospitals, and no equivalent dataset currently exists for the non-hospital-based outpatient sector (i.e., community-based outpatient care).

In the TOCCATA study, *PSA* data comprised diverse information on providers, patients, venue and billed services (e.g., hospital type or canton, patient sex, age, and nationality, service date, cost center, service codes, tariffs, and quantity)[g]. It was used to examine outpatient care patterns and continuity of care before, during, and after the COVID-19 lockdown. Each record corresponds to a unique outpatient encounter, enabling linkage with other datasets and longitudinal analyses of service use and costs.

Regarding data merging and access, *StatMed* and *PSA* data were linked over time at both the individual and hospital levels, then anonymized by the FSO before being provided to the research team under a fee-based contract.

3- *FKS dataset* (*Statistik diagnosebezogener Fallkosten, Statistique des coûts par cas, Statistica dei costi per caso*)

The *FKS* dataset offers a comprehensive overview of hospital case-related costs in Switzerland. Collected by the FSO since 1998, it serves as a key resource for healthcare cost analysis, hospital financial management, and policy evaluation. The dataset captures detailed information on the costs associated with individual hospital stays.

Each record includes variables related to patient characteristics—such as age, sex, place of origin, diagnoses, and procedures—aligned with the *StatMed* dataset. In addition, FKS offers a granular breakdown of hospital case costs, distinguishing between personnel costs and other operating expenses. This enables in-depth analyses of hospital resource allocation, cost structures, and financial dynamics at the case level.

In the TOCCATA study, *FKS* data were used to assess the economic impact of the COVID-19 crisis, particularly in relation to changes in cost patterns among cancer and non-cancer patients during and after the lockdown period.

4- Health insurance claims data for hospital and non-hospital outpatient care

[g] Referenzdaten - Allgemeine Richtlinien für alle Leistungserbringer-Typen: Forum Datenaustausch: Referenzdaten

The TOCCATA study also used anonymized claims data provided by two Swiss health insurers to examine outpatient services delivered outside of hospitals. These datasets complemented the *PSA* dataset by capturing services delivered in both hospital-based and non-hospital-based (community) outpatient settings[h], though they differed in the methods used to anonymize patient data and protect privacy.

One insurer supplied individual-level claims data for approximately 8,000 adult cancer patients and 20,000 randomly selected non-cancer controls who remained continuously insured throughout the study period. For each patient, the dataset included demographic information (e.g., age at service, sex, canton of treatment, date of death), insurance details (e.g., simplified model, deductible, cost-sharing), and comprehensive records of reimbursed services. These services included medications (ATC codes), hospitalizations, outpatient consultations (in-person and telemedicine), imaging (e.g., MRI, PET), oncology-related tests (e.g., biopsies, lab work), treatments (e.g., chemotherapy, radiotherapy), and non-medical care. Although the data originated from a single mandatory insurer, no additional selection bias was introduced during sampling.

The other insurer provided similar information but in a monthly aggregated format for both adult cancer and non-cancer patient samples. Following the enforcement of the new Federal Act on Data Protection (*nFADP*)[i], its Data Protection Officer (DPO) imposed a two- anonymity strategy (3-5) for several quasi-identifiers, including age class, sex, canton of residence, vital status, simplified insurance model, and simplified deductible. This anonymization process ensured that each combination of these variables appeared at least twice within the dataset, leading to a reduction of approximately 2,000 individuals from the sample initially selected based on inclusion and exclusion criteria. After anonymization, the final dataset included 18,856 non-cancer patients and 34,267 cancer patients, which is not representative of the overall population insured under mandatory health insurance.

Thus, insurance claims data from the first insurer were coded at the individual level, while those from the second insurer were anonymized and aggregated. Both datasets were provided to the research team under insurer-specific contracts established directly with the research team's institution. Importantly, PSA dataset and insurance claims data remained entirely separate and could not be linked because of legal, technical, and ethical constraints. For the same reasons, it was not possible to distinguish hospital-based outpatient services from community-based ones within the insurance claims data. As a result, it was not possible to reconstruct individual care trajectories across inpatient and outpatient settings, nor to assess the full continuum of care or the coordination between hospital-based and community-based services. Moreover, potential compensatory effects—such as increases in community-based care in response to reduced hospital-based outpatient services, or vice versa—could not be explored. These limitations restricted the ability to perform integrated analyses of outpatient service utilization, patient outcomes, and associated costs at the individual level.

Administrative health data general benefits and limitations

Administrative health data offers significant benefits for research and healthcare analysis, but it also comes with notable limitations (6).

Strengths of administrative health data

[h] Health insurance claims data encompass all outpatient care providers, including both hospitals and community-based services.

[i] Switzerland's new Federal Act on Data Protection (nFADP) (Nouvelle Loi sur la protection des données, nLPD) was adopted on September 25, 2020, and officially published in the Federal Gazette in December 2020. It became mandatory on September 1, 2023.

- **Large, population-wide coverage**: Administrative data often encompasses entire populations or large segments of the population, providing comprehensive insights into healthcare utilization, trends, and outcomes across diverse demographic groups.
- **Standardized coding**: With standardized coding systems (e.g., ICD for diagnoses, country-specific classification for procedures), administrative data ensures consistency and comparability across different settings, making it ideal for both cross-sectional and longitudinal studies.
- **Cost and utilization analysis**: Including detailed financial and service-use information, administrative data enables cost-effectiveness analyses, resource allocation assessments, and studies focused on healthcare access and equity.
- **Readily available and cost-effective**: Routinely collected for billing and administrative purposes, administrative data can often be accessed at a lower cost than primary data collection, reducing resource demands for research.
- **Longitudinal tracking**: Administrative data supports tracking of healthcare encounters over time, making it useful for studies on patients' trajectories, healthcare outcomes, and the long-term effects of healthcare policies.

Limitations of administrative health data

- **Limited clinical detail**: Primarily focused on billing and administrative needs, administrative data often lacks granular clinical information, such as lab results, imaging findings, or specific treatment outcomes, which can limit the depth of clinical insights and public health relevance.
- **Potential coding errors and variability**: Coding practices may vary across institutions, regions, and/or coders, leading to potential inaccuracies or inconsistencies in diagnoses and procedures. Over-coding or under-coding for reimbursement purposes can also impact data quality.
- **Lack of diagnostic detail in outpatient data**: In some healthcare systems, such as Switzerland, outpatient data may not include systematically coded diagnoses, making it challenging to analyze outpatient care adequately without supplementary sources.
- **Privacy and ethical considerations**: Due to the sensitive nature of personal health information, strict privacy and data protection regulations apply, often requiring de-identification and limiting access to detailed data.
- **Biases in data collection**: Since administrative health data is collected for billing rather than research purposes, it may reflect healthcare access limitations, reimbursement policies, or institutional priorities rather than actual patient needs or clinical priorities, potentially introducing biases.
- **Data fragmentation and lack of interoperability**: In many countries, insurance health data are held by numerous insurers, each potentially using different systems, standards, and formats. Accessing and standardizing data across multiple insurers adds complexity to data integration, interoperability, and comparative analysis in healthcare research and policy evaluation. In Switzerland, patients may change health insurers every year, which leads to a discontinuation of the documentation of longitudinal health care trajectories.

In summary, administrative health data is highly valuable for large-scale healthcare research and policy analysis due to its coverage and standardization. However, its limitations in clinical detail, potential biases, data fragmentation and privacy constraints require careful consideration and, often, supplementation with clinical data for more actionable and impactful results of the research.

Lessons learnt from the TOCCATA Project regarding administrative health data access

The TOCCATA project was initially planned for 12 months but was only completed after 3.5 years due to **prolonged delays in accessing administrative health data, particularly insurance claims data.** The challenges encountered throughout the project highlight critical regulatory, technical, and administrative barriers that impact health data utilization in Switzerland.

1. Challenges in accessing FSO Data

1.1. Contractual and legal constraints on *StatMed* and *PSA* Data

Accessing StatMed and PSA datasets required signing a "Contract for the One-Time Delivery and Use of De-Identified Individual Data in Switzerland" between FSO and Unisanté. Under the FSO's "one project, one request, one payment" policy, access was subject to the following conditions:

- Data could only be used once and exclusively within the defined scope of the research project.
- Data had to remain strictly anonymized, prohibiting any re-identification of individuals.
- Linking *StatMed* and *PSA* datasets with third-party or Unisanté-owned datasets was forbidden[j].
- Compliance with mandatory confidentiality protocols and security measures was required.
- Any data breaches had to be reported to the FSO within 72 hours.

These strict contractual and anonymization requirements limited research flexibility and prevented data linkage across different sources, thereby reducing the public health utility of the project.

1.2. Financial and logistical barriers

Unisanté was required to pay CHF 1,263 to cover FSO's data extraction and processing costs. While the initial data transfer occurred in March 2023, quality issues—such as missing links between *StatMed* and *PSA*—required further corrections, delaying the final datasets delivery until June 2024. These administrative bottlenecks prolonged research timelines and underscored the need for more efficient data-sharing mechanisms.

2. Challenges in accessing health insurance data

2.1. Data Access Agreements and legal constraints

To access insurance claims data, Unisanté researchers had to comply with Swiss data protection laws, including section 33 ATSG, the Swiss Data Protection Act (DSG), and its ordinances (DSV sections 1-3). Obligations included:

- Using the data exclusively for the specified project.
- Informing insurers of research findings.
- Securing data storage from unauthorized access.

[j] Under the Federal Statistics Act (FStatA) enacted on October 9, 1992., matching FSO data with third-party datasets is feasible but necessitates adherence to specific procedures. Certain conditions associated with this process, notably the requirement for the matching to occur within the FSO, may not be acceptable to external data holders, such as insurance companies. For more information, refer to the FSO website: Data linkages for third parties (Appariement de données pour le compte de tiers ;Datenverknüpfung für Dritte ;Collegamenti di dati per terzi)

- Prohibiting third-party sharing or cloud storage use.
- Complying with data protection laws and ensuring that anonymized data remains non-identifiable[k].
- Deleting, rendering unreadable, or returning the data within three months after project completion.
- Allowing the insurer and the insurer's Data Protection Officer (DPO) to review compliance to data protection measures as needed.

Failure to adhere to these obligations carried a CHF 10,000 penalty, payable to the insurer. This fine did not exempt researchers from their contractual duties but served as a financial deterrent.

2.2. Delays in data provision by insurers

Although the data-sharing agreement was signed in June 2023, one insurer required nearly an additional 15 months to deliver the requested data, which was ultimately transmitted in September 2024 following extended negotiations. The contract specified the following:

- Secure data transfer via an encrypted SFTP server within 15 business days of contract signing.
- Storage on Swiss servers with restricted access to designated personnel.
- Strict confidentiality clauses limiting third-party disclosure.
- Obligations to destroy or return confidential information at the project's conclusion.

However, the enforcement of the nFADP led to further restrictions on data anonymization, delaying access further (see above).

2.3. Impact of nFADP on sample size and data representativeness

Following the introduction of nFADP, one insurer strengthened data anonymization measures, affecting data access and usability. For instance,

- 2,000 patients were removed from the agreed dataset for privacy reasons.
- The modifications likely compromised the study population's representativeness and introduced biases that could not be quantified.
- The research team questioned the scientific validity of findings based on the altered dataset.

These restrictions emphasize the trade-off between privacy protection and research feasibility, necessitating a reassessment of anonymization strategies to ensure data usability and public health relevance.

3. Data linkage/matching challenges

3.1. Fragmented data sources and barriers to integration

[k] Re-identification of anonymized data can occur when quasi-identifiers such as age, sex, canton of residence, or medical history are cross-referenced with external datasets. Even if direct identifiers like names or social security numbers are removed, individuals can still be identified by combining seemingly innocuous pieces of information. For instance, rare medical conditions, unique treatment dates, or demographic characteristics can make a person identifiable within a dataset. The risk increases when anonymized data is linked to publicly available information, such as death records or insurance claims. In some cases, weak anonymization techniques—such as simple masking, pseudonymization without secure key destruction, or shuffling identifiers while retaining structured data—fail to prevent re-identification. Additionally, inference attacks can expose sensitive details, as patterns in anonymized datasets may reveal private medical information, even if individuals are not explicitly named.

StatMed and *PSA* data were linked longitudinally at the individual and hospital levels by the FSO, then anonymized before being made available to the research team under a paid cession contract. In parallel, insurance claims data—either coded (first insurer) or anonymized (second insurer)—were delivered under provider-specific agreements. These datasets remained entirely separate. Linking between hospital datasets (i.e., *StatMed* and *PSA*) with claims data from mandatory health insurances was not feasible due to legal, institutional, ethical, and practical constraints.

3.2. Legal frameworks restricting cross-source linkage

Beyond **the absence of a common patient identifier**, **Swiss regulations** impose stringent limitations on cross-source linkage to minimize the risk of re-identification. Relevant provisions include:

- The **Federal Statistics Act**, FStatA (*Bundesstatistikgesetz*, BStatG; *Loi sur la statistique fédérale,* LSF; *Legge sulla statistica federale*, LStat)[l], Art. 14a and its main associated ordinances:
  - Ordinance on the Matching of Statistical Data (*Datenverknüpfungsverordnung, Ordonnance sur l'appariement de données, Ordinanza sul collegamento di dati*)[m]
  - Ordinance on the Conduct of Statistical Surveys (*Statistikerhebungsverordnung, Ordonnance sur les relevés statistiques, Ordinanza sulle rilevazioni statistiche*)[n]
  - Ordinance on Fees and Compensation for Statistical Services of Federal Administrative Units (*Verordnung über die Gebühren und Entschädigungen für statistische Dienstleistungen von Verwaltungseinheiten des Bundes, Ordonnance sur les émoluments et indemnités perçus pour les prestations de services statistiques des unités administratives de la Confédération, Ordinanza sugli emolumenti e le indennità per le prestazioni di servizi statistici delle unità amministrative della Confederazione*)[o]
- **The FSO Data Linkage Directives**, version 1.23 published on March 05, 2020 (Verknüpfungsrichtlinien, Directives sur l'appariement)[p,q]
- **The Human Research Act**, HRA (*Humanforschungsgesetz*, HFG; *Loi relative à la recherche sur l'être humain*, LRH; *Legge sulla ricerca umana*, LRUm)[r], Art. 16(4) and its Human Research Ordinance, HRO (*Humanforschungsverordnung*, HFV; *Ordonnance relative à la recherche sur l'être humain,* ORH; *Ordinanza sulla ricerca umana*, ORUm)[s]
- **The revised/new Federal Act on Data Protection**, nFADP (*Neues Datenschutzgesetz*, revDSG; *Nouvelle loi sur la protection des données*, nLPD ; *Nuova legge sulla protezione dei*

[l] SR 431.01 - Federal Statistics Act of 9 October ... | Fedlex ; SR 431.01 - Bundesstatistikgesetz vom 9. Oktober... | Fedlex; RS 431.01 - Loi du 9 octobre 1992 sur la statist... | Fedlex ; RS 431.01 - Legge del 9 ottobre 1992 sulla stati... | Fedlex

[m] AS 2013 5475 - Verordnung des EDI über die Verkn... | Fedlex ; RO 2013 5475 - Ordonnance du DFI concernant l'ap... | Fedlex ; RU 2013 5475 - Ordinanza del DFI sul collegament... | Fedlex

[n] SR 431.012.1 - Verordnung vom 30. Juni 1993 über... | Fedlex; RS 431.012.1 - Ordonnance du 30 juin 1993 concer... | Fedlex ; RS 431.012.1 - Ordinanza del 30 giugno 1993 sull... | Fedlex

[o] SR 431.09 - Verordnung vom 25. Juni 2003 über di... | Fedlex ; RS 431.09 - Ordonnance du 25 juin 2003 sur les é... | Fedlex ; RS 431.09 - Ordinanza del 25 giugno 2003 sugli e... | Fedlex

[p] Verknüpfungsrichtlinien (Version 1.2; 05.03.2020) - | Publication, Directives sur l'appariement (version 1.2; 05.03.2020) - | Publication

[q] Data linkages for third parties ; Datenverknüpfung für Dritte ; Appariement de données pour le compte de tiers ; Collegamenti di dati per terzi

[r] SR 810.30 - Federal Act of 30 September 2011 on ... | Fedlex ; SR 810.30 - Bundesgesetz vom 30. September 2011 ... | Fedlex ; RS 810.30 - Loi fédérale du 30 septembre 2011 re... | Fedlex ; RS 810.30 - Legge federale del 30 settembre 2011... | Fedlex.

[s] SR 810.301 - Ordinance of 20 September 2013 on H... | Fedlex ; SR 810.301 - Verordnung vom 20. September 2013 ü... | Fedlex ; RS 810.301 - Ordonnance du 20 septembre 2013 rel... | Fedlex ; RS 810.301 - Ordinanza del 20 settembre 2013 con... | Fedlex

dati, nLPD)[t] and its implementing Data Protection Ordinance, DPO (*Verordnung zum Bundesgesetz über den Datenschutz*, VDSG; *Ordonnance relative à la loi fédérale sur la protection des données*, OLPD; *Ordinanza relativa alla legge federale sulla protezione dei dati*, OLPD), Art. 21.

To comply with these legal frameworks, the FSO anonymized the StatMed and PSA datasets prior to release. The insurers also applied their own anonymization or pseudonymization protocols before sharing data. As a result, there was no feasible way to directly link individual-level records across hospital and insurance sources.

3.3. Conditions for linkage under the FStatA, (BStatG, LSF, LStat)

The FStatA authorizes data linkage under strict conditions aimed at protecting individual privacy and ensuring the ethical use of data. First, linkages must serve exclusively statistical purposes and are explicitly prohibited for administrative or enforcement use. Second, a clearly defined legal basis or explicit consent from data subjects is required. Third, the law mandates transparency in data processing, alongside institutional accountability for compliance with data protection obligations.

While these provisions are fundamental for protecting privacy and maintaining public trust, they also introduce significant constraints on research. The exclusive focus on statistical purposes limits the scope of studies, particularly those aimed at operational decision-making or policy evaluation. In addition, the requirement for a legal basis or consent complicates project implementation and adds governance burdens. Strict anonymization protocols, while essential for confidentiality, can reduce the analytical utility of the data. Moreover, these constraints impose considerable administrative and technical demands on research teams and may discourage third-party data holders—such as insurers—from engaging in data linkage initiatives due to concerns over data sovereignty, confidentiality, and legal compliance.

3.4. Reinforcement by Federal and Cantonal Data Protection Laws

Swiss data protection legislation further reinforces the constraints on data linkage. The revised/new Federal Act on Data Protection (nFADP, revDSG, nLPD), which came into force in September 2023, imposes strict privacy requirements that explicitly prohibit the re-identification of individuals in research datasets. In parallel, both the HRA (HFG, LRH, LRUm) and cantonal data protection laws require that personal health data used for secondary research adhere to data protection principles, ensuring that any potential risks to individual privacy are mitigated[u].

In the Canton of Vaud, this requirement is codified in the Personal Data Protection Act, PDPA (*Loi sur la protection des données personnelles*, LPrD)[v], which regulates the processing of personal data by public institutions. In the Canton of Zurich, the law on Information and Data Protection Act, IDPA-ZH

[t] SR 235.1 - Federal Act of 25 September 2020 on D... | Fedlex ; SR 235.1 - Bundesgesetz vom 25. September 2020 ü... | Fedlex ; RS 235.1 - Loi fédérale du 25 septembre 2020 sur... | Fedlex ; RS 235.1 - Legge federale del 25 settembre 2020 ... | Fedlex

[u] In Switzerland, the reuse of non-genetic personal health data for research is governed by the Human Research Act (HRA) and the Human Research Ordinance (HRO) which provide a structured approach to the secondary use of non-genetic personal health data, particularly concerning data linkage:
- Explicit consent for identifiable non-genetic health data
- Presumed consent for coded data (i.e., individuals have been informed about the intended use and their right to dissent and consent may be presumed if no objection is raised)
- No consent for anonymized data which fall outside the scope of the HRA

When linking datasets, especially from different sources, it is crucial to ensure that the combined data do not lead to re-identification of individuals unless appropriate consent has been obtained. The process of data linkage must adhere to data protection principles, ensuring that any potential risks to individual privacy are mitigated.

[v] 172.65.pdf

(*Gesetz über die Information und den Datenschutz*, IDG)[w] follows similar principles, mandating lawful processing and anonymization of personal data when repurposed beyond its original intent. These cantonal regulations align with federal legislation by mandating that personal health data be anonymized to a degree that re-identification of individuals is not possible without disproportionate effort prior to any secondary use for research purposes.

Consequently, researchers must navigate complex legal considerations to ensure compliance, which may limit the granularity of data available for linkage and potentially impact the depth and scope of research findings.

3.5. Ethical obligations and the role of the HRA (HFG, LRH, LRUm)

Even when using fully anonymized data, ethical review obligations may still apply. Retrospective patient consent for linking hospital and insurance data is often infeasible, particularly at scale. Although this project posed minimal ethical risk due to the absence of identifiable data, the research team was nonetheless required to consult the relevant oversight bodies to determine whether ethics approval was required under the HRA (HFG, LRH, LRUm).

In practice, many cantonal ethics committees require formal notification or exemption requests, even for anonymized datasets. This process provides legal certainty and is essential for fulfilling publication standards, securing data access, and ensuring institutional compliance. Guidance from Swissethics[x]—the national coordination body for cantonal ethics committees—encourages researchers to seek confirmation of exemption even when no directly identifiable data are used.

3.6. Theoretical feasibility vs. real-world constraints

While the FSO has the legal and technical capacity to perform data linkage under the FStatA, such action requires the explicit agreement of all involved parties, including health insurers. In this project, such approval was not granted under the terms of existing data access agreements. Furthermore, navigating the complex authorization procedures—including legal, technical, and organizational steps—would have demanded considerable time and resources, incompatible with the research timeline and funding conditions.

Technical incompatibilities between hospital and insurance datasets further hindered linkage feasibility. Differences in coding systems, data formats, and documentation practices made integration challenging. These inconsistencies reduced the project's analytical potential and introduced risks of systematic bias, particularly the exclusion of patients with complex care trajectories. Such limitations have implications for equity and representativeness in health services research.

3.7. Fragmentation of governance and data standards

Institutional and governance fragmentation in Switzerland presents additional barriers to data linkage. The FSO enforces centralized rules and anonymization protocols, while insurers operate under decentralized models, each with its own data governance practices. This decentralization results in heterogeneity in anonymization procedures, update cycles, and data formatting. These discrepancies create substantial barriers to harmonization and interoperability, significantly complicating the construction of coherent, longitudinal datasets across care settings.

3.8. Resource constraints and administrative bottlenecks

[w] Gesetz über die Information und den Datenschutz (IDG) | Kanton Zürich

[x] swissethics.ch/en

Finally, operational limitations also delayed the project. Both insurers faced internal resource shortages, which affected the timeliness of data extraction and delivery. The complexity of anonymization procedures and protracted contract negotiations further contributed to delays. These challenges highlight the need for more standardized administrative workflows and institutional support mechanisms to facilitate timely and secure access to real-world data for research.

Lessons learnt from the TOCCATA Project regarding Swiss administrative health data usability

Several key challenges emerged regarding the use of administrative health data:

- **Lack of a data dictionary and metadata**: The data and variables provided often lack systematic explanation. Variables' construction, item values, and even variable names are unclear. This is especially true for TARMED insurance data, where variable names are open to interpretation and often lack transparency in construction decisions. Data providers should consistently include detailed notes on variable construction to retain essential context.
- **Inconsistent variable naming across data providers**: Variable names differ depending on the data provider (FSO, health insurers), complicating integration across sources. A standardized naming convention would greatly enhance data merging, as seen with the varying labels for "amount" across providers.
- **Lack of standardized grouping for variables**: Aggregated data often lack comparability. For instance, age groups may vary in aggregation (e.g., 0-4, 5-9 vs. 0-5, 6-10), making comparisons between sources challenging. Establishing grouping standards would be beneficial, at least within FSO datasets.
- **Consistency issues in *StatMed* data over time**: In *StatMed*, the last aggregated year does not include stays extending into the following year, creating a partial data gap. This issue can bias longitudinal analyses. Marking the last dataset year as provisional or delaying its closure to integrate more data would help mitigate this effect.
- **Large database volumes**: As data collection grows, so do storage and processing demand. For example, outpatient treatment databases (e.g., *PSA*) require significant computing power.
- **Coding variations in ICD-10 and CHOP**: Coding practices vary across hospitals, influenced by reimbursement-related motivations, leading to inconsistencies in patient information coding.
- **Gaps in health insurance data**: Health insurance claims data may not capture the full range of healthcare services used by patients, particularly those with high deductibles. In Switzerland, patients are responsible for paying healthcare costs out of pocket until they exceed their annual deductible. As a result, many individuals with high deductibles may choose not to submit smaller bills—such as those for general practitioner visits, minor procedures, or medications—because they know these expenses will not be reimbursed. Since health insurers only receive the invoices that are actually submitted for reimbursement, they may miss a significant portion of outpatient care delivered to these patients. This leads to gaps in the data and an incomplete picture of actual healthcare utilization, especially for preventive and routine services. These limitations can affect research and policy decisions that rely on insurance data to assess service use, healthcare needs, and costs across the population.
- **Absence of diagnoses in outpatient claims data**: In the Swiss healthcare system, outpatient claims data—whether from hospitals or health insurers—often lack systematically coded diagnoses. As a result, these datasets provide limited insight into the health conditions managed

in outpatient settings, making it difficult to conduct comprehensive analyses without the support of supplementary data sources.

- **Inability to link administrative health data across care settings:** The fragmentation of data sources creates a significant barrier to reconstruct individual care trajectories across inpatient and outpatient settings, and to assess the full continuum of care or the coordination between hospital-based and community-based services. This fragmentation arises primarily due to the separate reimbursement systems governing outpatient and inpatient care, as well as the absence of a unique patient identifier across datasets and legal constraints.
- **Loss of granularity in study results or analyses**: The level of anonymization required to comply with privacy regulatory framework often results in datasets being heavily aggregated or stripped of key pieces of information, reducing their usefulness for impactful research. To ensure that individual patients cannot be re-identified, sensitive variables such as age, residence, and rare conditions may be generalized or grouped into broad categories. While this approach enhances privacy protection, it reduces data granularity, making it harder to perform subgroup analyses or study specific population segments.

Recommendations to enhance administrative health data access and usability in Switzerland

To address the access and usability challenges of administrative health data, the following improvements are recommended:

- **Improving data documentation and standardization**
  - Providing a **comprehensive data dictionary** for each dataset, including metadata, variable definitions, and coding rules, would help users understand the data's structure and improve accuracy in data interpretation, facilitate data integration across different sources and ensure comparability over time. The FSO and health insurers could enhance transparency by offering these resources systematically across datasets.
  - **Harmonizing variable names** across data providers would facilitate data integration, making it easier to match/link datasets from different sources. This standardization would improve research efficiency and consistency, especially in studies involving multiple healthcare settings.
  - Standardizing age group categories and other aggregate metrics across datasets (e.g., hospital data vs. insurance claims data) would enhance comparability between sources, improving the quality of cross-sectional and longitudinal studies.
  - For datasets like *StatMed*, where year-end aggregation may exclude ongoing cases, clearly marking the final year as provisional or adjusting the cutoff date would minimize bias in longitudinal studies.
- **Streamlining data access procedures**
  - Implementing **more efficient and standardized mechanisms for researchers to request and obtain de-identified administrative data**, reducing delays, administrative bottlenecks and costs.
  - **Developing a regulatory framework for the secondary use of health data and for quality standards**: Establishing a regulatory framework for the secondary use of health data – encompassing access and processing within a Swiss Health Data Space- and setting clear

quality standards is essential to improve data reliability and usability. This framework should define quality benchmarks, such as data completeness, accuracy, timeliness, and consistency across datasets. It should also provide transparent guidelines for data access and governance, ensuring that all institutions and researchers adhere to, and operate under, common principles. This would improve the comparability and validity of research findings but also promote equity in access to data. By setting shared standards for quality, access and processing, the framework would also encourage data providers to maintain high-quality data, ultimately fostering greater trust in data-driven research, clinical decision-making, and policy development.

  - **Enhanced computing resources** for data management and analysis: Given the increasing size of datasets, researchers should access adequate secured computational resources, including high-performance computing for complex analyses. Additionally, fostering training in parallel computing and data optimization could improve data usability and access.

- **Enhancing data linkage capabilities**
  - Exploring **legal and technical solutions for privacy-preserving data linkage**, such as secure data enclaves or trusted third-party processing, to enable a more comprehensive analysis of patients' trajectories across health care settings.
  - Establishing a **Federal Health Data Trust Center** which would provide a secure infrastructure for linking various administrative and health data sources while protecting patient privacy. This center could serve as a secure, standardized data hub, enabling researchers to access linked datasets for longitudinal and multi-sectoral analyses. The Trust Center could operate under strict data governance policies and offer a secure remote access infrastructure (clean room) that maintain linkage capabilities for tracking patient journeys across different care settings and offer researchers to run their analyses for projects that are considered of public health relevance. This approach would streamline access, improve data quality, and enhance the ability to conduct comprehensive, impactful, privacy-compliant data-driven research in Switzerland.
  - Implementing a **privacy-compliant unique patient identifier** would facilitate more effective linkage between inpatient and outpatient datasets as well as between hospital-based and community-based outpatient datasets, enabling a comprehensive view of individual healthcare trajectories. Such linkage would be conducted by authorized experts within a trusted center, with researchers accessing only pseudonymized or anonymized datasets through a secure clean room environment. This approach would ensure both analytical value and compliance with data protection regulations.

- **Balancing privacy with research feasibility and utility**
  - Reassessing anonymization strategies to maintain a high level of privacy protection while preserving data granularity for meaningful analyses, particularly for vulnerable patient groups and rare conditions.
  - Setting clear guidelines on de-identification (anonymization; pseudonymization) processes that would allow researchers to understand data limitations and avoid potential biases introduced by anonymization, particularly when linking longitudinal data. This task could also be supervised or performed by a specific unit of the Federal Health Data Trust Center.

- **Strengthening interdisciplinary collaboration**

  - Encouraging partnerships between policymakers, researchers, healthcare providers, and data custodians to co-develop frameworks that promote responsible and effective use of administrative health data.

- **Exploring Electronic Patient Record (EPR) integration**: Transitioning from purely administrative data to incorporating EPRs (DEP, EPD, CIP)[y] would offer richer clinical insights, especially for studies focused on quality and outcomes of care as well as the ability to analyze health care trajectories. EPRs capture a broader range of patient information, including clinical notes, lab results, and treatment details, which can complement the focus of administrative data on billing and service records. Yet, for this to become efficient, a much larger proportion of the Swiss population should open an EPR.

Conclusions

This case study underscores the considerable potential of administrative health data to support health services research and inform public health responses, particularly in times of crisis. However, it also reveals key systemic barriers that currently hinder the efficient and effective use of these data in Switzerland and, therefore, highlights the difficulties to conduct impactful and timely public health relevant research.

Critical challenges include delays in data access, limitations in data linkage due to anonymization protocols and the absence of a unique patient identifier, and lack of interoperability across datasets. These issues reduce the timeliness, completeness, and comparability of data, ultimately affecting the reliability and generalizability of research findings.

To address these limitations, Switzerland must move toward a more integrated, standardized, and accessible health data ecosystem. Improving data documentation, harmonizing definitions, streamlining access procedures, and enhancing secure data linkage mechanisms are essential steps. Additionally, the integration of administrative datasets into Electronic Patient Records (EPRs) would enrich research with valuable clinical insights, enabling a more holistic evaluation of healthcare quality, patient trajectories, and treatment outcomes.

These efforts should be aligned with the objectives of the DigiSanté program and supported by strong interdisciplinary collaboration among researchers, policymakers, healthcare providers, and data custodians. Strengthening the governance and usability of administrative health data will be critical for enabling high-quality research, supporting evidence-based policymaking, and improving the resilience and responsiveness of the Swiss healthcare system.

References

1. Brunner L, Nicolet A, Peytremann-Bridevaux I, Marti J, Bulliard JL, Righi L, et al. Comparative inpatient care of cancer vs. non-cancer patients in Switzerland during the national COVID-19 lockdown: a nationwide interrupted time series analysis. BMC Cancer. 2025;25(1):477.
2. Tania A, OFS. Patients ambulatoires des hôpitaux en 2014: premiers résultats. Neuchâtel: Office fédéral de la statistique (BFS); 2016.

[y] See eHealth glossary: Dossier électronique du patient - DEP (Glossaire - eHealth Suisse : Définitions & explications), Elektronischen Patientendossier – EPD (Glossar - eHealth Suisse: Definitionen & Erklaerungen), Cartella informatizzata del paziente – CIP (Glossario – eHealth Suisse: definizioni e spiegazioni).

3. Sweeney L. K-anonymity: A model for protecting privacy. Int J Uncertain Fuzz. 2002;10(5):557-70.
4. El Emam K, Dankar FK, Issa R, Jonker E, Amyot D, Cogo E, et al. A globally optimal k-anonymity method for the de-identification of health data. Journal of the American Medical Informatics Association : JAMIA. 2009;16(5):670-82.
5. Lin WY, Yang DC, Wang JT. Privacy preserving data anonymization of spontaneous ADE reporting system dataset. BMC Med Inform Decis Mak. 2016;16 Suppl 1(Suppl 1):58.
6. Iezzoni LI. Coded data from administrative sources. In: Iezzoni LI, editor. Risk Adjustment for Measuring Healthcare Outcomes 4th ed. Chicago, IL: Health Administration Press; 2012. p. 95-146.