Automating Health Care Data Validations

Lessons from Colorado's All Payer Claims Database (CO APCD) APM, DR, and VBPC Programs

Dagmar Velez, MPH, MS
Lindsay Wilkins, MPH
Center for Improving Value in Health Care

Background

All payer claims database (APCD) programs depend on accurate, non-claims-based submissions to support transparency and policy decisions, directly affecting the reliability of insights used to guide health care spending, regulation and system performance. Annual files such as Alternative Payment Models (APM), Drug Rebate (DR), and Value-Based Pharmaceutical Contracting (VBPC) contain complex datasets submitted by payers, often including three to four years of historical data.

Each submission requires extensive validation across multiple dimensions, including comparisons to prior submissions, APCD data, and year-over-year reporting. Historically, this validation process has been manual, time-intensive, and difficult to scale, creating significant burden for analysts and increasing the risk of inconsistencies.

Objective

To design and implement an automated validation framework that:

- ☑ **Reduces manual effort and processing time**
- ☑ **Improves accuracy and consistency of validation outputs**
- ☑ **Standardizes validation logic and reporting language**
- ☑ **Enables scalable, repeatable validation across payers, file types, and reporting years**

Operational Improvements

Before the addition of automation, the validation process for the Annual Files was time-extensive, complex and highly repetitive. Implementation of a Python-based automation solution drove substantial operational improvements:

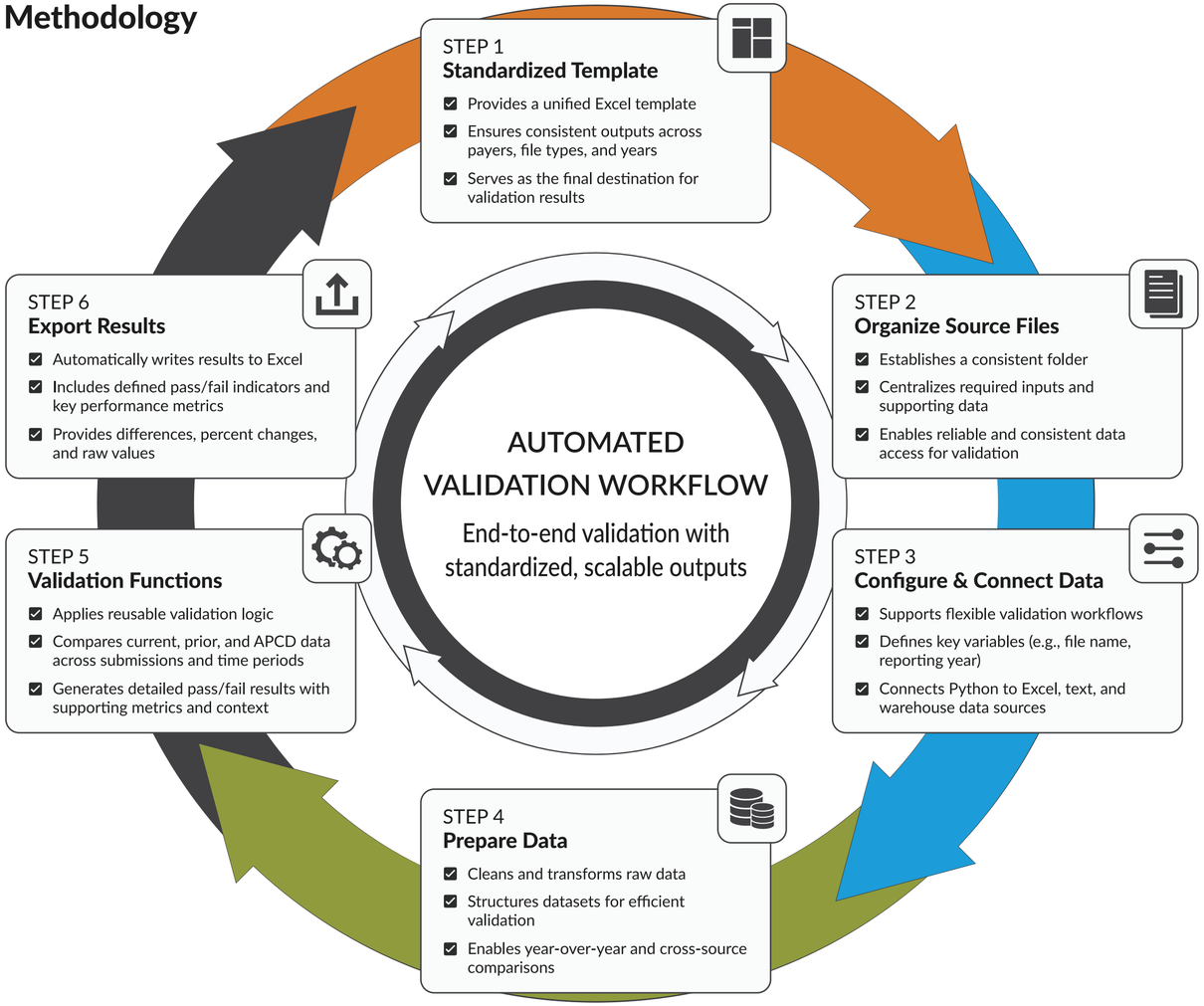
50% Time Reduction
Validation time cut from 200+ to ~90 hours annually

Scales to 20+ Payers
Handles multiple versions + diverse file types

Standardized Validation
Consistent logic replaces manual variability

Automated Reporting
Excel outputs with pass/fail + detailed metrics

Methodology



Impact

Implementation of a Python-based automation process significantly improved efficiency and data quality.

Improved Data Quality
Automation reduced manual entry and review errors, resulting in more consistent, reliable submissions.

More Actionable Outputs
Enhanced validation logic produced clear, detailed and structured validation results, enabling faster identification of data issues and more efficient resolution by submitters.

Increased Analyst Efficiency
Automation reduced repetitive, manual tasks, allowing analysts to focus more time on higher-value activities such as more in-depth analysis and insight generation.

Conclusion

Automating the validation of non-claims-based health care data submissions substantially improves efficiency, accuracy, and scalability. Replacing manual SQL-driven workflows with a flexible Python-based framework reduced analyst burden, minimized errors, and produced more consistent and actionable outputs.

Beyond supporting current validation needs, this approach creates a scalable foundation for future enhancements, including more advanced analytics, expanded validation logic, and continuous improvement in health care data quality.

Although implementation details may vary across APCDs and organizations, this framework demonstrates that validation automation can be successfully adapted to different reporting environments and provides a practical model that others can build from.

Additional Information

For more detailed information on how to create your own Python-based automation, please scan the QR code below.
For questions: info@civhc.org

Scan to view the automation workflow presentation

Learn more about automated health care data validations at civhc.org

CIVHC
CENTER FOR IMPROVING VALUE IN HEALTH CARE