Consiglio Nazionale delle Ricerche

IRCRES

ISTITUTO di RICERCA sulla CRESCITA ECONOMICA SOSTENIBILE
RESEARCH INSTITUTE on SUSTAINABLE ECONOMIC GROWTH

Counterfactual evaluation of Artificial Intelligence solutions for Industry: Which way forward?

Sella Lisa, *Ragazzi Elena, Benati Igor*
(lisa.sella@ircres.cnr.it)

Special Session S43
Counterfactual methods for regional policy evaluation

Presentation outline

- CH4I: the project and some notions
- Methodological approaches to CH4I impact evaluation
- CH4I case studies (description, evaluation design)
  - ✓ Healthcare system
  - ✓ Agrifood system
- Discussion

CH4I Circular Health for Industry — The project

- *Circular Approach to Health*: whole **ecosystem** where humans live, i.e., strong **interconnections** among urban, rural, wild ecosystems (Capua, 2020)
- *Strong demand for digital transformation* in **healthcare** and **agrifood** systems (traceability, certification, normative compliance, evidence-based approaches)
- *Need for sharing data* among ecosystems (fragmentation problem)
- *Artificial Intelligence* and other digital technologies can create a **paradigm shift**

- *CH4I mission*: Developing methodologies and **infrastructures** for collecting data to create **digital twins** and **machine learning algorithms**
  - ✓ *Case 1*: Improve the *use of resources* and the *quality* of **healthcare services**
  - ✓ *Case 2*: Improve *food quality and safety, animal welfare and productivity* in **agrifood industries**

IRCES

CH4I

A multiform partnership

Fondazione Compagnia di San Paolo

di.unito.it
DIPARTIMENTO DI INFORMATICA

FONDAZIONE BRUNO KESSLER

AGRINNOVA

DISAFA

AZIENDA OSPEDALIERO - UNIVERSITARIA
Città della Salute e della Scienza di Torino

Caritas Christi urget nos!
OSPEDALE COTTOLENGO

FUDEX

Vanzetti
HOLSTEIN

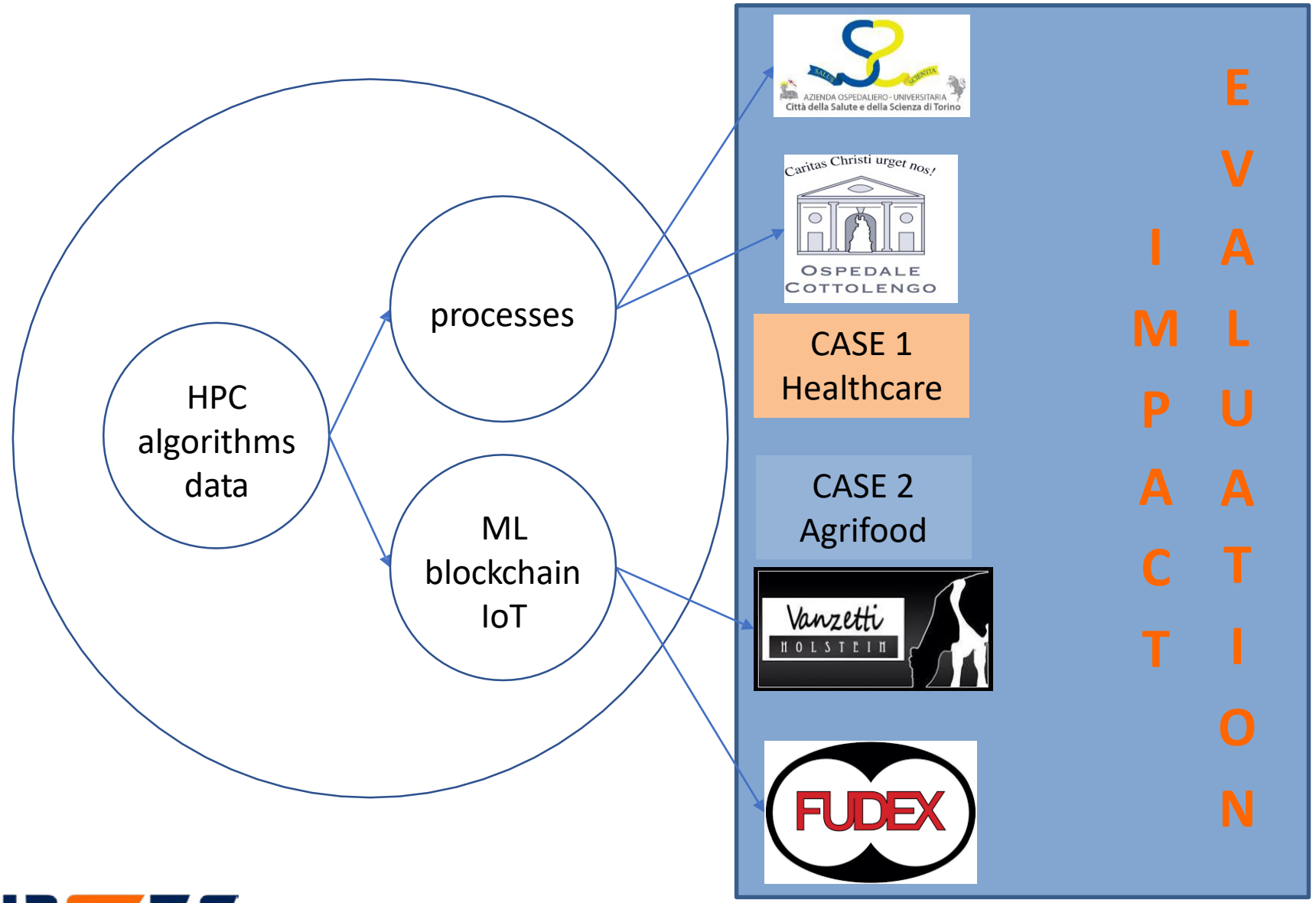
+CIM4.0

consoft SISTEMI

Consiglio Nazionale delle Ricerche
IRCRES
ISTITUTO di RICERCA sulla CRESCITA ECONOMICA SOSTENIBILE
RESEARCH INSTITUTE on SUSTAINABLE ECONOMIC GROWTH

CH4I Circular Health for Industry

The project



Evaluating CH4I Methodological approaches

- *Logical modelling*: fixing the **chain of events** linking CH4I actions to their results and (potential) impacts
- *Counterfactual impact evaluation*: applied to CH4I **actions implemented in the real world** (Oakden-Rayner & Palmer, 2019; D'Alberto et al. 2018; Gentler et al., 2016)

  ***Research question*:** Are KPIs better due to CH4I actions?

  Estimation of the expected impact
- *Counterfactual reasoning*: **applied to Digital Twins**, when policies are not implemented in real workflows (Barricelli et al., 2020; Alber et al., 2019; Molnar, 2019)

  ***Research question*:** How could real KPIs get better if the suggested policies (process innovations) are implemented in real workflows?

  Upper bound of the expected impact (?)

IRCES

CH4I case 1 — Healthcare system

- *Motivation*: the provision of **high quality hospital services** has an impact on patients' life quality (and not only)

  But they depend on an **efficient execution of processes** in **highly uncertain** contexts

- *AI 4 healthcare*: manage and **adapt healthcare processes as fast** as possible to circumstances (optimization of procedures, scheduling, organizational aspects)

  Process mining: applied to *clinical and administrative processes*

  process discovery – automatic extraction of models by event logs (recording)

  real-time process management – online optimization with lookahead

  predictive process monitoring applied to ongoing process execution

  Digital Twin: digital representation of hospitals/wards

  detection of inefficiencies by *workflow analysis*

Circular Health for Industry

IRCES

Hospitalization at Home (HaH): alternative hospitalization method addressed to some categories of disease/patient (uncomplicated ischemic stroke, congestive heart failure, dementia from behavioral disorders, etc.)

about 500 patients per year

Actual criticalities of HaH service provision:

a. Some suitable patients never selected;

b. Some unsuitable patients selected;

c. Sometimes long indirect ways across wards, exacerbating health risks

AI Objectives: **optimizing patients' fluxes** from Emergency Room (ER) to HaH

- AI algorithm assisting ER MDs to immediately identify suitable HaH patients (i.e., HaH real-time alert)
- Improve HaH success/undertaking
  a. Improving direct flux from ER to HaH (suitable patients)
  b. Reducing ways back from HaH to ER (not suitable patients)
  c. Shortening ways to HaH (indirect ways)

CH4I case 1.1

AZIENDA OSPEDALIERO-UNIVERSITARIA
Città della Salute e della Scienza di Torino

HaH Impact Evaluation

First evaluation proposal

- *RCT experiment on HaH real-time alert*:

the AI algorithm identifies all eligible patients, but the alert to ER MDs is randomized

Research question: Does the alert increase *pr(HaH|eligible)*? And consequent impacts on pathways length (shorter waiting time in ER), health risks, healthcare costs

***Criticalities*:** healthcare system stickiness, **the alert cannot be implemented in practice**

Second (feasible) evaluation design

- *Counterfactual reasoning*: on retrospective observational data

Research question: What if the HaH alert is operationalized? (Simulation data)

Real outcome vs. Simulated outcome -- Potential effects on KPIs:

- number of HaH patients
- waiting times
- resources (sustainability)

***Criticalities*:** this is a **potential impact** (human decision-making in real processes)

IRCES

CH4I case 1.2

AZIENDA OSPEDALIERO - UNIVERSITARIA
Città della Salute e della Scienza di Torino

Interventional Radiology

Diagnostic and therapeutic procedures that are minimally invasive and based on medical imaging guidance – 1 operating theatre, about 40 patients per week

Actual IR criticalities: a. operating theatre *underuse*; b. staff *overrunning and overtime*; c. large amount of time dedicated to *scheduling*

AI Objectives: optimizing organizational procedures and use of IR resources (equipment and staff)

- AI algorithm assisting IR MDs in real-time scheduling (Digital Twin)
- Improving IR efficiency and reducing costs

  a. Reducing waiting times (i.e., dead times; real-time prediction of operation duration and delays)

  b. Reducing staff overtime

  c. Reducing staff involvement in scheduling (work-time saving)

  d. Better foreseeing and managing of critical situations

IRCES

CH4I case 1.2

AZIENDA OSPEDALIERO-UNIVERSITARIA
Città della Salute e della Scienza di Torino

IR Impact Evaluation

First evaluation stage

- *Counterfactual reasoning on IR Digital Twin*: retrospective observational data

***Research question*:** What if scheduling is totally based on IR Digital Twin? (simulation data)

Real outcome vs. Simulated outcome -- Potential effects on KPIs:

- number of IR operations
- waiting times
- overtime (i.e., staff costs)

Second evaluation stage

- *Cluster randomization experiment*: MD-assisted scheduling during randomized weeks (treatment); counterfactual weeks: standard MD scheduling

***Research question*:** Does the MD-assisted scheduling improve KPIs of interest?

Criticality: spillover effect on MDs decision-making (same staff in factual and counterfactual weeks)

IRCES

CH4I case 2 Agrifood system

- *Motivation*: important economic sector (17% of Italian GDP), **vulnerable** to challenges posed by ***agricultural practices*** in the field and by ***contamination*** throughout the processing chain
- *Objectives*: improving food quality and safety; strengthening the resilience of food processing
- *AI 4 agrifood*

  Cloud-based, sensors, drone technologies monitoring:

  - soil texture, moisture, water stress, crop status, pests in the field
  - contaminants throughout the processing chain

Circular Health for Industry

IRCES

CH4I case 2.1 Healthy Crops for Breeding

Data collected through *atmospheric stations and drones* to improve the **quantity and quality of crops** for cattle breeding in a partner farm (Vanzetti Holstein)

Objectives:

a. production of healthy crops for breeding

b. positive impact on cattle's health

AI techniques **identify optimal actions** in plant cultivation:

- Optimal and timely *irrigation*
- Interventions to *prevent infections* and plant fungi in crops
- Optimal timing for *sowing and harvesting*

CH4I case 2.1 HCfB evaluation

Impact evaluation analysis conducted on KPIs related to the quality and quantity of plant production on **three different soils**

Research question: Does the use of AI techniques produce healthier crops for breeding?

Criticalities:

a. small number of fields

b. limited availability of historical data

c. we cannot exclude spillover effects

CH4I case 2.2 Safe Food Production

Data collected through *sensors* positioned on **packaging lines** of a partner firm (Fudex)

Objectives: improve **quality and safety** of packaged food products

AI techniques to limit (avoid) the presence of *heavy metals and other potential pollutants* inside packages:

- improve the quality of *controls*
- optimize the *production process*
- avoid *waste* of product

SFP Evaluation

Comparison of **quality and safety KPIs** between traditional packaging line and packaging line with AI systems

Production line No AI

Production line With AI techniques

Research question: Does the use of AI techniques have an impact on quality and safety of food products?

Criticalities:

a. preserve the same functioning of the two packaging lines
b. unreliable historical data
c. professional skills of packaging line staff

Conclusive remarks & criticalities

- **Actual need for AI applications** to real problems; but AI solutions are very **specific**
- While adapting the general research question *(do AI approaches improve quality, efficiency, sustainability of industries?)* to practical contexts, we assisted to a **proliferation of case studies** (i.e., dispersion of resources)
- **Difficult implementation of organizational innovations in real contexts** (stickiness of real processes, staff opposition to innovations, bureaucratic barriers)
- **Difficult implementation of a counterfactual evaluation approach** (cultural and organizational barriers). In most case studies, the *time horizon* is too short to implement experimental AI applications in real environment. Difficult evaluation of true impacts.
- The funding institution demanded for an impact evaluation design, but its commitment was really modest and it was not able to impose the adoption of a true counterfactual approach … **This is the cultural challenge evaluators face at the moment: CIE is demanded for by funding agencies, but they give no empowerment to the evaluators to implement it**

Thanks for your attention

Sella Lisa	**lisa.sella@ircres.cnr.it**
Benati Igor	igor.benati@ircres.cnr.it
Ragazzi Elena	elena.ragazzi@ircres.cnr.it

https://ch4i.di.unito.it/

References

[1] Barricelli, B. R., Casiraghi, E., Gliozzo, J., Petrini, A., & Valtolina, S. (2020). Human Digital Twin for Fitness Management. *IEEE Access*, *8*, 26637-26664.

[2] C. Molnar. (2019). *Interpretable Machine Learning. A Guide for Making Black Box Models Explainable*. [Online]. Available: ttps://christophm.github.io/interpretable-ml-book/

[3] S. Wachter, B. Mittelstadt, and C. Russell, ``Counterfactual explanations without opening the black box: Automated decisions and the GDPR,'' *Harvard J. Law Technol.*, vol. 31, p. 841, 2017.

[4] Alber, M., Tepole, A. B., Cannon, W. R., De, S., Dura-Bernal, S., Garikipati, K., ... & Kuhl, E. (2019). Integrating machine learning and multiscale modeling—perspectives, challenges, and opportunities in the biological, biomedical, and behavioral sciences. *NPJ digital medicine*, *2*(1), 1-11.

[5] Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B., & Vermeersch, C. M. (2016). *Impact evaluation in practice*. The World Bank.

[6] Oakden-Rayner L., Palmer L.J. (2019) Artificial Intelligence in Medicine: Validation and Study Design. In: Ranschaert E., Morozov S., Algra P. (eds) Artificial Intelligence in Medical Imaging. Springer, Cham. https://doi.org/10.1007/978-3-319-94878-2_8

[7] Chen, X., Li, W. L., Zhang, Y. L., Wu, Q., Guo, Y. M., & Bai, Z. L. (2010). Meta-analysis of quantitative diffusion-weighted MR imaging in the differential diagnosis of breast lesions. *BMC cancer*, *10*(1), 693.

[8] Scarvell, J. M., Galvin, C. R., Perriman, D. M., Lynch, J. T., & van Deursen, R. W. (2018). Kinematics of knees with osteoarthritis show reduced lateral femoral roll-back and maintain an adducted position. A systematic review of research using medical imaging. *Journal of biomechanics*, *75*, 108-122.

[9] D'Alberto, R.; Zavalloni, M.; Raggi, M.; Viaggi, D. AES Impact Evaluation With Integrated Farm Data: Combining Statistical Matching and Propensity Score Matching. *Sustainability* 2018, *10*, 4320.

[10] Rubin, D.B. Estimating causal effects of treatments in randomized and nonrandomized studies. J. Educ. Psychol. 1974, 66, 688–701.

[11] Pufahl, A.; Weiss, C.R. Evaluating the effects of farm programmes: Results from Propensity Score Matching. Eur. Rev. Agric. Econ. 2009, 36, 79–101